VALIDATING TESTS FOR PERSONNEL SELECTION

BY CYRIL BURT

I. *Statistical problems in testing for the Forces* (pp. 1–2). II. *Predicting efficiency in mechanical work: Tests and testees* (pp. 3–4). III. *Methods of weighting with a composite battery* (pp. 4–7): (*a*) *Weighting by correlation coefficients* (p. 4); (*b*) *Weighting by partial regression coefficients* (p. 5); (*c*) *The unweighted sum of standardized marks* (p. 5); (*d*) *Tests of significance* (p. 6). IV. *Methods of correcting for effects of selection* (pp. 7–13): (*a*) *Proportional weights in the unselected population* (p. 7); (*b*) *Proof of selection formulae* (p. 8): (1) *Multivariate selection* (p. 9), (2) *Univariate selection* (p. 10); (*c*) *Illustrations of selection formulae* (p. 10): (1) *Univariate selection* (p. 10), (2) *Multivariate selection* (p. 11); (*d*) *Estimating the probable number of misfits* (p. 12). V. *Empirical verification of selection formulae* (pp. 13–17): (*a*) *With inter-correlations not completely given* (p. 14); (*b*) *With inter-correlations completely given* (p. 16). VI. *Practical corollaries* (pp. 17–18). VII. *Summary* (pp. 18–19). *References* (p. 19).

I. STATISTICAL PROBLEMS IN TESTING FOR THE FORCES

The task of validating tests for use in the Army and other war-time services has raised several problems in psychological statistics that must first be solved before we can safely gauge the efficacy of the various tests proposed. When plans for such inquiries were first put forward for discussion, I ventured to suggest, in memoranda illustrated by results obtained from tests in the Army and elsewhere (14, 15, 16), that most of the problems, though wearing a slightly different guise, were in all essentials the same as those that confronted the educational psychologist in the early stages of his work, and that the present requirements might therefore be met by taking over, no doubt with minor adaptations, solutions long ago worked out or adopted in the field of educational testing. The subsequent progress of investigations, in the Army, Navy and Air Force, would seem to confirm both the need and the feasibility of the methods suggested.

Accordingly, it may be of assistance to the numerous workers in this and allied fields if the most useful devices are described in a more accessible form. The non-statistical investigator will rightly be sceptical of improved results obtained by means of mere arithmetical calculation; and one of my chief objects will therefore be to test the main formulae on data which will enable the deductions to be checked by first-hand observations.

In what follows, to save constant circumlocution, I shall use the word 'test' to cover, not only marks or measurements reached by formal testing, but also gradings, ratings, and assessments for any kind of physical or mental trait, however obtained. The first problem, then, is that of combining a battery of several 'tests' in order to secure a single assessment for a man's suitability either for a particular trade or for promotion to higher rank. Here various procedures have been suggested. In theory the best is undoubtedly to calculate a set of partial regression coefficients, and weight each test accordingly; but for practical purposes, I hope to show that, provided arbitrary differences in the marking scales have been eliminated, a straight sum or a simple average may often give almost as good a result.

The second problem is this. It is manifestly impracticable to apply special tests for each arm or branch to every new recruit. Is it then possible, either by factor analysis or by job analysis, to reduce the chief qualifications to a small set of factors, so that, at any rate for first preliminary allocations, one composite examination may serve for all? If we regard a 'factor', not as a unitary trait or ability, but as a weighted pattern, including

character qualities and types of past education or experience, as well as natural aptitudes, then the same set of 'tests', differently weighted, might conceivably be made to predict efficiency in several divergent directions.

The third problem is more easily overlooked. It arises from the difference in relative heterogeneity between the kind of group on which tests have usually to be validated and the much wider group for which the tests are eventually required—a difference that may at once reduce and distort the relevant correlations.

This last issue will best be understood if we consider a concrete example. Let me take a question put to me soon after the outbreak of the war by an officer commanding an anti-aircraft battery which I was visiting: How would a psychologist set about discovering what qualities are most needed by men who are to be chosen for training as gun-layers? The scheme generally adopted for such an inquiry is sufficiently familiar. Having devised tentative tests or other means for measuring what seem to be the relevant qualities, the psychologist proceeds to compare the results thus obtained at the beginning of the training period with some independent assessment for each man's efficiency in the actual job, if possible after the training is over; in this way he hopes to ascertain the relative value of each quality tested, and the accuracy with which ultimate efficiency could have been predicted from the initial measurements alone. But a first-hand assessment of actual efficiency can only be secured for those who have already been selected and trained. Consequently, it can only tell us that, of the members of a given team, A has proved more efficient than either B or C; it could not tell us that X and Y (in the Pioneers, it may be) would have been more efficient still, and ought to have been sent to the battery instead of B and C. The correlations taken just as they stand, therefore, may easily mislead us in regard to the true merits of the tests as a whole, and the relative merits of each one. For instance, with certain tests (employed in 1940, but now discarded) we found that the correlation for an intake trained at a local anti-aircraft depot amounted to only 0·36 for tests of a 'verbal' type and 0·25 for those of a more 'practical' type(14). I shall endeavour to show that the low value of correlations such as these (and of others like them which have often been cited in this connexion) can be largely explained by the narrow range of variability that is almost inevitable within a selected group. Further, since selection is likely to have reduced variability in practical abilities most of all, the higher correlation obtained with the non-practical or 'verbal' tests may be due simply to a special peculiarity of the selected group, not to any superior reliability or suitability in the tests themselves. The whole difficulty is rather like the time-honoured fallacy of judging the efficiency of a scholarship examination as a means of *selection* by stating its efficiency as a means of predicting the order of merit *within* the selected group (cf. (4), p. 64, (5), p. 206).

There are other incidental problems calling equally for a statistical solution. Is there, for example, any quick method of item analysis, which will enable us to throw out the least useful items from each composite test? What is the best way of estimating correlations when the criterion is given, not in a graded, but in a qualitative form (e.g. 'suitable' and 'unsuitable')? How can we best determine the 'pass-mark' that is to divide the suitable from the unsuitable? These, too, are questions which psychologists had to face in the early days of practical work, when their chief sphere of activity lay in the educational rather than in the vocational field, and with children rather than with adults (cf. e.g., (5), pp. 165, 202, 220). Without at least some provisional solution of such problems, it would have been quite impossible to standardize mental and scholastic tests for allocating children to appropriate types of school—secondary, technical, mentally defective, and the like; and a fairly long experience of the working of the devices then proposed has confirmed their practical utility.

The advantage of trying out new formulae with children first of all is evident enough. A psychologist working in the elementary schools of a wide area, like London, can obtain large and relatively unselected samples; he can apply a great variety of tests, entailing far more time than adult examinees can spare; and he can procure from competent and experienced observers, namely, teachers already trained in methods of mental assessment, highly reliable judgements with which to compare and check his tests. At the present moment, it would obviously be undesirable to publish illustrative inquiries based on the testing now going forward among the Forces themselves. I shall, therefore, endeavour to exemplify the procedure proposed by means of earlier data. Every one of the methods described below has been applied to actual test results obtained in the Army, Navy, or other services; but it will, I think, be found that the material here substituted yields in the end a more convincing verification than could have been hoped for, had we taken data for men who have only recently been tested, and for whom only meagre reports are at present available for confirming or comparing the efficacy of the tests.

II. PREDICTING EFFICIENCY IN MECHANICAL WORK: TESTS AND TESTEES

In a war of mechanized armies, one of the most useful services the psychologist can render is to standardize tests for picking out suitable personnel for what we may call 'mechanical' occupations, whether in the factories or in the field. On a smaller scale and with younger people, Education Departments in industrial areas have long been faced with a somewhat similar problem. In London, for example, many inquiries were made by the Council's psychologist, working in collaboration with teachers, examiners, and other officials, on the best procedure for sorting out lads best fitted for training in the Council's trade schools and technical institutes with a view to placing them later on with engineering firms. It was found that the problem was inevitably bound up with the selection of boys and girls for somewhat different purposes—for scholarships at secondary schools and for transference to central schools with a commercial or an industrial bias.

Preliminary investigations ((4), pp. 58–9) indicated that, so far as abilities (as distinct from character traits) were concerned, the initial requirements both of educational and vocational selection could be reasonably met, if the first classifications were based on tests of four main 'factors', general intelligence, verbal ability, numerical ability, and mechanical ability respectively (g, v, n, k). It is instructive to note that much the same four factors have come to the front in analyses of tests applied to men and women in the Forces. All these so-called 'abilities' are undoubtedly complex; and the last has given rise to a series of tests that have been labelled with different names according to the problems or the prepossessions of the investigators—'practical', 'mechanical', 'manual', 'kinaesthetic', 'spatial' and even 'aestheto-kinetic' tests. For trade schools (as they were originally termed) the commonest procedure has been to set a written examination in English subjects and arithmetic, and to supplement this with more practical devices for assessing the pupils' ability in various kinds of 'mechanical' work. The relative value of the different parts of such examinations has been continually disputed: some have declared that the written part of the examination is entirely irrelevant; others have urged that verbal or non-verbal tests of intelligence should be introduced in addition or instead; others again have claimed that, particularly for the so-called mechanical abilities, school reports based on actual work and progress would be far more trustworthy.

To gain light on these issues a number of researches were carried out, often extending over a number of years. Standardized tests were given year after year to boys entering (or likely to enter) certain trade schools; and an effort was made to compare the results both with the lads' subsequent progress in the trade school and with reports secured later on after each had remained for some time with an engineering firm. For those entering branches of the engineering industry, the tests chiefly employed were group tests of (1) intelligence (verbal and non-verbal), (2) English (silent reading, spelling, and composition), (3) arithmetic (including both mechanical and problem work), and individual tests of (4) handwork (in various forms). The last three were largely adaptations of the ordinary examinations conducted for trade scholarships and entrance to technical schools.[1]

[1] For assistance at many different points, especially in the construction and standardization of the more technical tests, we were greatly indebted to various head and assistant masters, and to investigators working with or for the National Institute of Industrial Psychology. In particular, I received much personal help from Dr Ritchie Scott, Principal of the Beaufoy Institute, and Mr Max Tagg, Principal of the Acton Polytechnic.

I shall deal first with the predictive efficiency of the tests. Unfortunately, to obtain satisfactory data after the training period was over proved far more difficult than to obtain plausible assessments at the outset of that period. Hence the numbers available are comparatively small. However, over a period of years it proved possible to collect, for a fair proportion of those tested, gradings, more or less trustworthy, for workshop efficiency from the firms they eventually entered. The gradings were checked, supplemented, and equated by similar gradings obtained from the trade school itself just before the lads left; and the pooled results provide the best obtainable 'criterion' by which to gauge the merits of the several types of test. But our present problem centres in the value of the methods rather than in the value of the results. Consequently the possible inadequacies of the criterion need not concern us here: a fuller discussion of the details has been given in earlier reports: (cf. also this *Journal*, XIV, pp. 343 *et seq.*).

III. Methods of weighting with a composite battery

The number of lads in the composite group was 354; and at the time of testing the majority were between the ages of $12\frac{1}{2}$ and $13\frac{1}{2}$. With this group the correlations of the four tests with each other and with the criterion are shown in Table 1. From the correlations with subsequent trade efficiency we might be at once tempted to infer that tests and examinations of a verbal or linguistic type are far less important than tests of a more practical type and particularly a test of general intelligence. This, indeed, seems amply confirmed by later researches, published and unpublished. And the question consequently arises: How can we best weight each of the tests (if we decide to retain it), so as to achieve the best possible assessment of subsequent efficiency?

Table 1. *Observed figures for the selected group*

Test	S.D.	Correlations					Multiple correlation
		Intelligence	English	Arithmetic	Handwork	Total	
Intelligence	0·765	(1·000)	0·688	0·579	0·433	2·700	—
English	0·782	0·688	(1·000)	0·617	0·334	2·639	—
Arithmetic	1·034	0·579	0·617	(1·000)	0·362	2·558	—
Handwork	1·013	0·433	0·334	0·362	(1·000)	2·129	—
Total	—	2·700	2·639	2·558	2·129	10·026	—
Subsequent trade efficiency	0·953	0·545	0·292	0·346	0·494	1·677	0·625
Regression coefficients	—	0·513	−0·208	0·059	0·319	—	—

(*a*) *Weighting by correlation coefficients*

In earlier investigations it was a common practice for the writers to suggest that each test should be weighted according to its correlation with the criterion; and this simple suggestion has recently been revived for investigations on military tests. In general such a procedure is defensible only as a rough and ready device. Under certain rare conditions, namely, when the inverse of the correlation matrix is approximately equivalent to a unit matrix, it happens to be correct. But primarily the correlation with the criterion merely expresses the value of each test *in isolation*: when the test is combined with others in a battery, the proper weighting is that furnished by its partial regression coefficient. With the foregoing figures, were we to take the criterion correlations as weights, we should reach a multiple correlation of 0·562. The difference between this value and that obtained by using the partial regression coefficients is scarcely significant with a group of this size; but in certain cases and with larger groups the method might give a very unjust impression of the efficacy of the test battery.

The reluctance of research students in this country to adopt the method of partial regression seems partly due to the notion that the arithmetic involved is long and difficult; but partly (if we may judge by discussions embodied in several research theses) it is the result of a lengthy controversy provoked by the use of such coefficients in one or two early L.C.C. Reports.[1] In answer to the theoretical criticisms then raised, it will here perhaps be sufficient to remark that in other branches of applied science (e.g. economics and agriculture) the method of partial regression has been increasingly employed by statisticians of every school, though no doubt with increasing caution. As to the practical objections, a little experience with the computations required will show that the labour need not be serious if the number of tests is small, and a simplified arithmetical procedure is adopted: the working method described in the Appendix to my original memorandum—'single division' to obtain one row of regressions, followed by a 'back solution' for the remaining rows—will be found both simpler and speedier than that suggested by Yule (which British students mostly use), or by Guilford (Doolittle's method), or by Thomson (Aitken's method of pivotal condensation).

(*b*) *Weighting by partial regression coefficients*

For calculating partial regressions the formula employed is, in matrix notation,

$$\beta = R^{-1}r, \qquad \text{(i)}$$

where β and r denote single columns, consisting of the regression coefficients and the criterion correlations respectively, and R^{-1} is the inverse of the matrix of inter-test correlations ((17), p. 297, eq. 264, (10), p. 150). The weights or partial regressions so calculated ('beta coefficients') are shown in the last row of figures in Table 1 (p. 4). These put the tests in a somewhat different light. We now discover that, if an intelligence test is used, the English test, when included at all, must be given not a positive but a negative weight.

To estimate the improved correlation between the criterion and the joint result of the tests when weighted by these figures, we apply the well-known formula for multiple correlation ((7), p. 300):

$$r_{0.1\ldots n} = (\beta' r)^{\frac{1}{2}}, \quad \text{i.e.} \quad \sqrt{(\Sigma \beta_{0i} r_{i0})}. \qquad \text{(ii)}$$

It is necessary to emphasize that in this equation the criterion correlations are weighted with the *standardized* regression coefficients ('beta coefficients'), not with the *crude* ('*b* coefficients') as has been done in at least one recent published research. With the present selected group, the multiple correlation amounts to 0·625. This figure is decidedly higher than that obtained by taking even the best single test (intelligence) by itself. The regressions, however, suggest that the positive correlation of both English and arithmetic with the criterion is due mainly (though in the case of arithmetic not entirely) to their correlation with intelligence; if both these subjects are omitted, the multiple correlation remains almost as high as before, viz. 0·616.

(*c*) *The unweighted sum of standardized marks*

In an earlier memorandum ((11), p. 304), I pointed out that much of the improvement in the net correlation, attributed by investigators to the weighting of the tests with their respective regression coefficients, really arises from the elimination of the arbitrary and irrelevant weighting unconsciously introduced when unstandardized marks are simply added just as they are. Unless every test and every examiner allot marks on much the same scale and with much the same effective range, the ensuing differences in the standard deviations of themselves amount to differential weighting. Accordingly, by merely reducing the marks to standard measure (or better, making their standard deviations roughly proportional to the partial regressions) the chief disturbing influences may thus be removed, without any need for the complicated arithmetic that an explicit regression equation entails for each case.

To determine the correlation of the criterion with the unweighted sum or average of the standardized

[1] I had ventured to urge that, for research purposes, a method of measuring intelligence, more accurate than the Binet scale, would be to take a composite set of tests, each yielding graded measurements, and to weight each one according to the partial regression coefficients ((3), pp. 140 f.). This and other uses of such coefficients ((5), p. 183) gave rise to emphatic comments on their limitations (see, for instance, (8), and a series of articles by Wechsler, Thomson, Burke and others in the same *Journal*). I have elsewhere pointed out that, in theory at any rate, the limitations attending partial regressions also affect the formulae for correcting for selection to be described in a moment.

marks, the formula required is similar to that used in calculating a factor-saturation[1] by 'simple' as distinct from 'weighted' summation, viz.,

$$r_{0(1+\ldots+n)}=\frac{\Sigma_i r_{i0}}{\sqrt{(\Sigma\Sigma_{ij} r_{ij})}}, \qquad \text{(iii)}$$

where 0 as usual denotes the criterion, and i and j any of the n tests. From the totals in Table 1 we have at once

$$r_{0(1+\ldots+n)}=\frac{1{\cdot}677}{\sqrt{(10{\cdot}026)}}=0{\cdot}524.$$

The difference is statistically significant. In this case, therefore, the unweighted average of the standardized marks seems to be far from giving as good a prediction as the weighted; but, as we shall see in a moment, the result is impaired, not only by the inclusion of the English test, with a positive weight, but also by the way the group tested has been selected.

Although the tests of English and arithmetic here seem to add little to the value of the battery, we are not to conclude that these tests will be always useless. In most of the examinations for trade schools, indeed in nearly all such examinations at the time this inquiry was originally planned, there was no test of intelligence. Let us then drop intelligence from our computations, and recalculate the partial regressions. We now find that the weighting for the English test becomes positive, and that for the arithmetic test is appreciably increased. The conclusion is plain. When no intelligence test is included, the tests of English and arithmetic form a valuable addition to the test of handwork: the three tests must then be weighted in the proportions of about 1, 3, and 7; and we can expect a correlation with the criterion of about 0·52. However, a far better procedure would be to use an intelligence test in conjunction with a test of handwork, and weight them in the proportions of about 4 to 3: the regressions are then 0·407 and 0·318 respectively, which yields the multiple correlation cited above, 0·616. But when an intelligence test has been included, the tests of English and arithmetic become all but useless—the English test, indeed, worse than useless. Thus the common practice of giving these two tests equal weight with handwork, or (in some cases) even more, means that we are choosing pupils for the engineering trades largely on account of scholastic aptitudes or skill that are really irrelevant to such work.

(*d*) *Tests of significance*

For determining the standard errors of the partial regressions and multiple correlation, the formulae given by Kelley ((7), p. 302, eqs. 282, 283) will, as a rule, be sufficiently accurate. Strictly speaking, the denominator in his formulae should be based on the number of degrees of freedom, not on the total number in the sample (N). It will be seen that the standard error of the multiple correlation is virtually identical with that of an ordinary correlation: thus, with the figures given in Table 1, it is

$$\frac{1-r^2_{0.12\ldots n}}{\sqrt{(N-n-1)}}=\frac{1-0{\cdot}625^2}{\sqrt{(354-5)}}=0{\cdot}032.$$

Since, however, the sampling distributions do not tend to the normal, the significance of a borderline case should be tested by an analysis of variance ((10), p. 247). For the standard error of the partial regression coefficient ($\beta_{0i.12\ldots n}$ say), Kelley's version of the equation is not so convenient as the determinantal ((12), p. 305): viz.

$$\sigma_\beta=\sqrt{\left\{r^{ii}\left[\frac{1-r^2_{0.12\ldots n}}{N-n-1}\right]\right\}}. \qquad \text{(iv)}$$

To test the significance of the difference between two regressions, allowance must be made for the fact that they are not independent; the formula is therefore

$$\sigma_{\text{diff.}}=\sqrt{\left\{[r^{ii}+r^{jj}-2r^{ij}]\left[\frac{1-r^2_{0.12\ldots n}}{N-n-1}\right]\right\}}. \qquad \text{(v)}$$

(The notation r^{ij}, as usual, denotes the element in the ith row and jth column of the inverse of the matrix

[1] Cf. (11), p. 284, eq. xxv. A further formula there given for unweighted correlations (eq. xli) also holds good of weighted; and will be found useful to determine how many tests should be included in a battery in order to achieve a multiple correlation of a desirable size, say r_{kg} on the assumption that the average inter-test correlation is $\bar{r}_{kk}$. It yields a practical corollary: unless the inter-correlations are much smaller than the correlations with the criterion, to increase the number of tests beyond about half a dozen at most is likely to be waste of time.

of inter-test correlations.) For instance, taking the regressions for intelligence (0·513) and for handwork (0·319), we find

$$\sigma_{\text{diff.}} = \sqrt{\left\{[2{\cdot}195+1{\cdot}260-2\,(-0{\cdot}476)]\left[\frac{1-0{\cdot}625^2}{354-5}\right]\right\}} = 0{\cdot}087:$$

the difference is thus fully significant.

IV. METHODS OF CORRECTING FOR EFFECTS OF SELECTION

The data on which we have so far based our conclusions relate only to a selected, and indeed a relatively homogeneous, sample of the school population, namely, lads from trade schools. The weights just computed, therefore, might enable us to predict how trade school pupils will turn out later on, when compared with each other; but they will not tell us how to use the tests for the prior and more important purpose of sifting the general population for the most suitable trade school pupils. For this broader problem, we require, not the correlations found among those who have already entered the trade schools, but the correlations found among the entire, unselected mass of school children generally. Unfortunately, as we have seen, it is generally impossible to extend first-hand observations for the criterion beyond the selected group. In consequence, it becomes urgently desirable to have some means of 'correcting' the correlations and the weights found in the narrower sample, and so inferring the probable correlations and the appropriate weights for the population as a whole. The methods and proofs outlined below were described more fully in an L.C.C. Memorandum on *Correlations obtained from Selected Groups* (1934), and in a later Memorandum on *Weighting Tests for Personnel Selection*(14): the Appendix to the latter describes the working methods in detail, and can be obtained from the Psychological Laboratory, University College, London.

(*a*) *Proportional weights in the unselected population*

In weighting a battery of tests for practical purposes, all we need to know are the *proportionate* values of the regression coefficients. To estimate these for the general population the only additional information required is some indication of the degree to which selection has reduced the heterogeneity of whatever traits we are testing within the smaller and narrower sample. Such a measure we can obtain by simply taking the ratio of the standard deviation in the sample to the standard deviation in the population at large. This ratio, $h_i = \sigma_i/\Sigma_i$, I have termed the 'coefficient of heterogeneity for the sample'. The *reduction* in heterogeneity will then be measured by $\sqrt{(1-h_i^2)} = H_i$; and this I call the 'coefficient of homogeneity'. To obtain the proportionate values of the weights the formula is then

$$u_{01.2\ldots n} = \beta_{01.2\ldots n}/h_1 = \frac{\Sigma_1}{\sigma_1}\beta_{01.2\ldots n}, \qquad \text{(vi)}$$

where $u_{01.2\ldots n}$ denotes the weighting coefficient for test 1 in the *general* population and $\beta_{01.2\ldots n}$ the partial regression coefficient obtained in the selected *sample*. To reduce the weights to fractions of unity for purposes of comparison, it is convenient to divide them throughout by their numerical sum, $\sum_i |u_{0i.j\ldots n}|$. With these formulae (i and vi) there is no need to calculate first of all the various partial correlations, and then deduce the corresponding partial regressions from these.

As part of our inquiries into the distribution and relations of educational abilities, all the tests described above were applied to children in ordinary elementary schools as well

as to boys in trade schools. Accordingly, we possess full data for their variabilities and inter-correlations in the general population from which the trade school lads have been drawn. But we have no similar data either (i) for the standard deviation of the criterion or (ii) for the correlation of each test with the criterion within that wider population. (During the later stages of the survey, in preparation for the part that local education authorities seemed likely to play in the choice of employment for children under their charge, teachers in certain elementary schools were invited to assess, so far as they could, the probable suitability of their pupils for trade schools and mechanical or technical occupations; and these assessments will be used later for a rough check on the results. We shall, however, see that the teachers' assessments for trade suitability could claim nothing like the accuracy of their assessments for work of a more academic type; nor are the groups so covered altogether comparable with those for whom first-hand assessments of actual efficiency in engineering firms could eventually be obtained. I shall therefore keep to this latter type of criterion first of all, although the sample is smaller.)

The figures for the standard deviations in the selected group and in the general population are given in Tables 1 and 2. To render comparisons simpler, I have taken the standard deviation of the intelligence test as unity. The proportionate weights for both groups are shown at the foot of Table 2. Here, it will be seen, the corrections in proportionate weights prove after all to be comparatively slight. Unless, however, we actually calculate the effects of selection, we can never be sure that the weights have not been seriously affected. The largest change is in the weight for intelligence. Here, paradoxically enough, the selection has *increased* the figure for the selected sample. That is because selection was largely based on other (and more specific) qualities, closely allied to the criterion. In such cases we may easily be tempted to over-estimate the importance of intelligence tests for the purposes of the selective examination, unless we guard against such a source of error by making an actual computation.

Table 2. *Estimated figures for the general population*

(Figures in brackets are data, not estimates)

Test	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Intelligence	(1·00)	—	0·832	0·777	0·695	—
English	(1·24)	0·832	—	0·841	0·708	—
Arithmetic	(1·58)	0·777	0·841	—	0·711	—
Handwork	(1·47)	0·695	0·708	0·711	—	—
Subsequent trade efficiency	1·103	0·684	0·550	0·577	0·659	—
Regressions	—	0·578	−0·279	0·079	0·399	0·742
Proportionate weights:						
(*a*) General population	—	0·433	−0·209	0·059	0·299	—
(*b*) Selected sample	—	0·469	−0·186	0·054	0·291	—

(*b*) *Proof of selection formulae*

In any theoretical investigation, however, we shall want to know how large a multiple correlation with the criterion we are to expect with these tests and with these weights in the general population; and this cannot be estimated unless we know, not merely the proportionate, but also the absolute values, of the corrected weights.

Formulae which are applicable for this purpose were devised by Karl Pearson in connexion with a somewhat different problem ((1), p. 17, eqs. xlv, xlvi, xlvii). But both his

proof and his equations can be greatly simplified. The issue is one of many with which, as I have argued elsewhere, the psychologist can cope best if working with covariances and variances rather than correlations. In spite of the objections urged against this proposal by Dr Stephenson, it is hard to see how a difficulty like the present could be solved without either tacitly or explicitly adopting such functions. In any case, even if accepted, his criticisms would only invalidate the use of absolute measures of variance, not a statement of their relative values, which alone are here in question.

(1) *Multivariate selection.* Let us use the capital letter, C, to denote any matrix or submatrix of variances and covariances in what I may call the *given* group, i.e. the group for which all the relevant data are known; and the next letter, D, to denote the corresponding matrix or submatrix in the *derived* group, i.e. the group for which certain constants have to be deduced. (In Pearson's proof, the 'given group' is always the larger or parent population, and the 'derived group' a smaller and more homogeneous sample obtained from the former by selection. In validating educational or vocational tests the 'given group' will usually be the smaller and more homogeneous sample: our task is then to work backwards, and deduce figures for the larger and more heterogeneous population from which the selected sample has been drawn.)

Let the subscript a stand for those tests or traits for which we have direct and complete information in regard to the effects of selection (that is to say, we know their variances and covariances in both the given group and the derived group on which the selection has been based); and let the subscript x stand for the remainder. Here, as in most correlational problems, when considering the applicability of the formulae, and the way the results are to be interpreted, it will be wise to examine the causal framework implied by the data, and to make sure that it is consistent with the tacit assumptions on which the proof of the formulae has been founded. The simplest type of causation is that in which the variables designated a are those on which the selection has been directly based, while the variables designated x are only affected indirectly. I shall therefore refer to the former as the 'selective' variables, and the latter as the 'non-selective'. Thus, to take the foregoing inquiry as an example, we may suppose (at any rate for the sake of argument) that the youths in the trade schools have been directly selected from the general population on the basis of assessments or examinations for Intelligence, English, Arithmetic and Handwork, which were sufficiently similar to the tests that have supplied our own data for the two to be treated as identical: on the other hand, we may regard the variability for 'trade efficiency' as indirectly affected only. (In practice the distinction will be merely relative: and, unlike other writers on the subject, I hold that the real distinction is a distinction in knowledge, not in causation.)

Accordingly, C_{aa} will designate the 4×4 submatrix of variances and covariances, for the four tests mentioned, within the selected sample; D_{aa} will denote the corresponding submatrix obtained from the general population. Our data include actual figures for (i) both these matrices; they likewise include figures for (ii) the variance of the assessments of trade suitability and (iii) the covariances between those assessments and the four tests, among the trade school sample, i.e. for the matrices C_{xx} and C_{ax}: the corresponding matrices for the general population, namely, D_{xx} and D_{ax}, are what we have to deduce.

To reach a simple algebraic solution let us begin with the common assumptions (i) that any regression with which we are dealing can be sufficiently represented by a linear function, and (ii) that the linear function adopted will be that which yields the closest possible fit as judged by the principle of least squares. Keeping to the usual notation, let $w_{x1.2\ldots n} \equiv \beta_{01.2\ldots n}\, \sigma_0/\sigma_1$ denote the partial regression of the criterion 0 on test 1 (unstandardized): we may call it the 'crude weighting coefficient' for test 1. Then, in matrix notation, and with the same convention as to letter, we may write $W'_{xa} = D'_{xa} D^{-1}_{aa}$ to denote the row vectors of all such weighting coefficients in the derived population, and $V'_{xa} = C'_{xa} C^{-1}_{aa}$ to denote the corresponding row vectors for the given population. It is easy to see, without formal proof, that (apart from sampling errors) any selection of children in regard to the selective traits (those designated a) should not affect the rectilinear regression of any non-selective trait, x, on the several selective traits, a: i.e. $W'_{xa} = V'_{xa}$. Hence $D'_{xa} D^{-1}_{aa} = C'_{xa} C^{-1}_{aa}$. Consequently, to calculate the criterion covariances in the larger population from the figures for a given selected sample, the equation required will be

$$D'_{xa} = C'_{xa} C^{-1}_{aa} D_{aa} = V'_{xa} D_{aa}. \qquad \text{(vii)}$$

Again, let $(X - X_A)$ be the specific elements in X remaining after A has been eliminated: and let F and G denote the matrices of 'partial' variances and covariances, after the elimination of A, in the given

sample and the general population respectively. Then $F=G$, since they must be unaffected by selection in regard to A. Further, since by definition $(X-X_A)\,X'_A=0$,

$$F=(X-X_A)\,(X-X_A)'=(X-X_A)\,X'=XX'-V'_{xa}AX'=C_{xx}-V'_{xa}C_{ax}.$$

Hence, $$D_{xx}-V'_{xa}D_{ax}=G=F=C_{xx}-V_{x'a}C_{ax};$$

and $$D_{xx}=C_{xx}+V'_{xa}\,(D_{ax}-C_{ax}), \qquad \text{(viii)}$$

and $$C_{xx}=D_{xx}-W'_{xa}\,(D_{ax}-C_{ax}). \qquad \text{(ix)}$$

Though derived from somewhat different premises and expressed in a somewhat different form, the formulae we have thus obtained are virtually those originally deduced by Karl Pearson on the more limited hypothesis that in both the unselected *and the selected population* the variables are distributed in accordance with the normal law. By employing generating functions along the lines suggested by Soper(6), Aitken has since supplied a short and elegant proof in matrix form(9). However, for the ordinary student of psychology the arguments of all these writers are difficult to follow. Moreover, I venture to think that, though algebraically equivalent, the version here reached is not only much simpler to understand, but much easier to apply. Aitken's methods are admirably illustrated in a recent study by Emmett(18).

(2) *Univariate selection.* The matrix equations just proved are generalized formulae for multivariate selection. If, however, the selection is based on one variable only, the matrix equations resolve themselves into simpler forms which can readily be expressed by the more usual notation. In accordance with the usual practice, let capital letters (Σ, R) denote values for the larger population and small letters (σ, r) those for the selected sample. Then, $C_{aa}=\sigma_a^2$, $D_{aa}=\Sigma_a^2$, and the covariance for test x in the sample$=c_{ax}=r_{ax}\sigma_a\sigma_x$: so that

$$R_{ax}=\frac{\sigma_x}{\Sigma_x}\cdot\frac{\Sigma_a}{\sigma_a}\,r_{ax}, \qquad \text{(x)}$$

or, writing $\sigma/\Sigma=h_x$ (as above), $$R_{ax}=\frac{h_x}{h_a}\,r_{ax} \quad \text{and} \quad r_{ax}=\frac{h_a}{h_x}\,R_{ax}. \qquad \text{(xi)}$$

Similarly, writing $H=\sqrt{(1-h^2)}$ (as above), equation (viii) can also be expressed in a simple form; and we have for any two tests, x and y,

$$r_{xy}=\frac{R_{xy}-H_xH_y}{\sqrt{(1-H_x^2)}\,\sqrt{(1-H_y^2)}}, \qquad \text{(xii)}$$

which (as Karl Pearson has shown) is a more general version of Yule's well-known formula for partial correlation. From this we obtain

$$R_{xy}=r_{xy}h_xh_y+\sqrt{(1-h_x^2)}\,\sqrt{(1-h_y^2)}. \qquad \text{(xiii)}$$

These simplified expressions are the familiar formulae for 'univariate selection' (cf. Kelley(7), pp. 225, 229). As I have elsewhere shown, they may also be directly deduced from Yule's well-known equations for partial correlation: we have

$$\frac{r_{xy}-r_{ax}r_{ay}}{\sqrt{(1-r_{ax}^2)}\,\sqrt{(1-r_{ay}^2)}}=r_{xy.a}=R_{xy.a}=\frac{R_{xy}-R_{ax}R_{ay}}{\sqrt{(1-R_{ax}^2)}\,\sqrt{(1-R_{ay}^2)}}, \qquad \text{(xiv)}$$

$$\sigma_x\sqrt{(1-r_{ax}^2)}=\sigma_{x.a}=\Sigma_{x.a}=\Sigma_x\sqrt{(1-R_{ax}^2)}, \qquad \text{(xv)}$$

$$r_{ax}\frac{\sigma_x}{\sigma_a}=b_{xa}=B_{xa}=R_{ax}\frac{\Sigma_x}{\Sigma_a}. \qquad \text{(xvi)}$$

And the above formulae follow at once by inserting in (xiv) the values for $r_{ax}r_{ay}$, h_a, and h_x given by (xv) and (xvi) respectively. The initial matrix equations set out in the previous proof (p. 9) are in fact merely generalizations of these premises, extended to cover several tests instead of one.

More detailed discussions of these various formulae, with working methods, and some account of their uses and limitations in the field of education, were given in the memoranda cited above. Here my object is chiefly to urge that they have equal value for researches on vocational guidance, especially those of the type now contemplated in the fighting forces. Accordingly, I shall first illustrate the results to be obtained in the cases of univariate and multivariate selection respectively, and then offer some empirical evidence for the trustworthiness of the figures deduced by these equations.

(c) *Illustrations of selection formulae*

(1) *Univariate selection.* As noted above, much the same tests have been applied to large and relatively unselected samples of the general school population, though no one large or fully representative group has taken *all* the four tests. We thus possess fairly

trustworthy figures for the variances within the general population, but no good estimates for the inter-correlations. As before, we may take the ratio of the standard deviations of the tests in the selected trade school group and in the general school population respectively as indicating how far heterogeneity has been reduced in each case by selection on the basis of presumable trade efficiency; apart from these measurable changes in variability, there appear to be no grounds for suspecting that any important factors, outside the scope of our data, have seriously affected the inter-correlations within the selected group. Hence we may justifiably attempt to estimate the inter-correlations for the general population by means of the equation for the univariate selection given above (xiii).

The estimates so calculated are shown in Table 2. The *actual* correlations for the general population, as computed in some of my earlier L.C.C. Reports (e.g. (4), pp. 52 f.), were, as a matter of fact, somewhat lower. But the coefficients there given were obtained by taking 'complete age groups' in certain elementary schools only. Thus the age groups were *not* really 'complete' in the broader sense of including, not merely all children of a given age in the particular schools tested, but also a due proportion of brighter children from better neighbourhoods and from central and secondary schools and of duller children from poorer neighbourhoods and from schools for the mentally defective. For the present investigation, in order to demonstrate beyond all doubt the inevitable effects of selection, it seemed better to take figures from a 'general population' that was as complete and as heterogeneous as possible. Moreover, as is shown by the coefficients published in later reports, the accuracy of scholastic testing has steadily improved. When allowance has been made for these small differences, the estimated figures and the actual agree quite well.

(2) *Multivariate selection.* To complete our inquiry, and illustrate the complex calculations involved in cases of multivariate selection, let us accept the variances and covariances between the four tests both for the selected sample (calculated from Table 1) and for the general population (calculated from Table 2) as our known data; and from these let us deduce the probable values of the correlations between the tests and the criterion in the general population. I should observe that in a well-planned investigation it would not be desirable to make two steps or stages in estimating the influence of selection on the final figures; but our primary purpose here is, as I have already insisted, not so much to deduce valid results, as to illustrate the different methods available; and space is saved by using the same set of figures to exemplify the two sets of formulae.

For our present problem we require first of all to estimate $D_{xx}=\Sigma_x^2$, the variance of trade efficiency as it would appear in the entire youthful population, supposing the whole of that population could first be trained in the technical schools. To find this figure we must employ the second multivariate equation (viii). Performing the necessary matrix multiplications and subtractions, we finally reach $\Sigma_x^2=1{\cdot}217$, and therefore $\Sigma_x=1{\cdot}103$. With the aid of the first multivariate equation (vii) we can deduce the criterion covariances in the general population from the corresponding figures for the selected group; and, using the standard deviation just determined, we can then compute the required correlations and the absolute values of the partial regression coefficients within the general population; or we can deduce them direct from the sample regressions (see fifth and sixth lines of Table 2). By their means we can verify the figures for the proportionate weightings given in the seventh line of the same table, and already calculated by the shorter method first described. On comparing the coefficients, reached by these correction formulae, with the uncorrected figures directly obtained from the selected sample and printed at the

outset in Table 1, it will be seen that the process of selection has changed the criterion correlations far more than it has changed the weights.

In a theoretical study like the present, we naturally desire some estimate of the multiple correlation to be expected in the general population: this is the figure by which our battery of tests must be judged, if we are to use them for the prior purpose of selection, as distinct from that of differentiating among individuals already selected. It is for this purpose, as we have seen, that we need the absolute values of the regression coefficients. Moreover, the new figures for the correlations plainly warn us that, unless we are content with the common practice of under-estimating the selective efficacy of the composite battery, by taking the observed correlations from the selected sample just as they stand, it is imperative to make proper allowance for the effects of selection.

Taking equation (ii) and inserting the new figures, we find that the multiple correlation for the general population works out at $r_{0.1234} = 0{\cdot}742$. This value is appreciably larger than that computed from the observed correlations, which was, it will be remembered, only 0·625; and the difference is statistically significant. If, instead of weighting each test, we are content to use the simple sum or average of the standardized marks, then the correlation of this sum with the criterion can now be estimated for the general population by using equation (iii) and inserting the new values for the inter-test correlations (Table 2). The multiple correlation obtained in this way proves to be 0·682; the difference between this figure and that obtained by weighting is here not statistically significant. If we omitted the test of English (which has a negative regression), then the multiple correlation obtained with the simple sum or average would be higher still, namely, 0·707. On turning back to the corresponding figures, computed before we have corrected for selection, we shall see that they might easily have misled us on these points.

(*d*) *Estimating the probable number of misfits*

It remains to translate the theoretical coefficient of correlation into some more practical figure which will enable us to compare the amount of wastage entailed with and without the proposed scheme of tests or examinations. What the administrator wants to know is this: If we allocate our candidates at random, then we shall inevitably be faced with a large number of misfits; but suppose we introduce a scheme of psychological testing: how far will that succeed in reducing the probable number of misfits? To answer this question we can use either Pearson's Tables ((2), II, pp. 79–137) or what I have called 'tetrachoric graphs' (cf. (5), p. 219); but now, instead of employing given percentages for failures and successes to deduce the correlation, we employ the correlation to deduce the number of failures and successes.

An example will make the procedure clear. One of the commonest borderlines, both in educational and in other work, is drawn in the neighbourhood of the 10th percentile (1·28 S.D.). Some educational authorities, for instance, offer free places at secondary and central schools to the best 10 % of the school population (or thereabouts), and have proposed to establish special classes or schools for the most backward 10 %. Similarly, in the fighting services we are often asked to pick out the most efficient 10 % for special training or to eliminate the most incompetent 10 %. The calculations given above in the earlier sections indicate the kind of multiple correlation we may reasonably assume with a battery of 4 to 6 tests—namely, about 0·70. (In the Army the observed coefficients are far lower, but require correction both for the unreliability of the criterion and the

selective character of the samples available.) Taking the figures just cited by way of illustration, we discover from the graph or table that, if we pick out with our tests the best 100 men from an intake of 1000, we can expect only 47 of them to be included in the 100 picked out by the criterion (practical efficiency in the field); as many as 53 will not be included. It may seem surprising to learn that over half of those who pass with the criterion fail on the tests. But two points must be remembered. First, had we picked out our 100 men at random, nearly double the number, namely, 90, would probably have been below the standard imposed by the criterion. Secondly, with the tests, those who *do* fall below that standard will not fall *far* below it: many will be borderline cases.

Accordingly, particularly as the criterion is itself so unreliable, it would seem better to adopt a different form of stating 'wastage', and inquire how many of the 100 selected would be above or below *average* efficiency, as judged by the criterion. Adopting this procedure, we find that only 5 out of the 100 picked out by the tests would be below average efficiency as judged by the criterion; whereas as many as 50 would have been below average, had they been selected at random.

It should be noted that, with normal correlation, in order to state the number of misfits or 'wastage', a different graph is necessary for every different borderline. If, however, we are content with a simple 'point' correlation and with *both* borderlines drawn through the average or medians, there is a very simple relation: the amount of 'semi-wastage' (w) can be estimated by the formula $s-w=\frac{1}{4}(1-r_p)$. Thus, with a point correlation of $r_p=0{\cdot}50$, the amount of semi-wastage is 12·5 %: i.e. if, out of a total intake of 100, 50 are above average on the criterion, only about $12\frac{1}{2}$ of those 50 will be missed by the tests; and another $12\frac{1}{2}$, selected as above average by the tests, will be below average according to the criterion. It must be remembered, however, that the normal correlation, r_t, is much higher than the point correlation, r_p (i.e. the product-moment correlation based on the assumption that the four frequencies are all concentrated at a single point). In fact, $r_t=\sin\frac{1}{2}\pi r_p$. Thus, if $r_p=0{\cdot}50$, $r_t=0{\cdot}70$ (cf. (16), pp. 174, 28).

V. EMPIRICAL VERIFICATION OF SELECTION FORMULAE

It is now clear that, by using formulae to correct for selection, we may arrive at estimated coefficients for the general population that are far higher than those directly calculated from the selected sample. But can we trust these armchair magnifications? The objection is one that is constantly raised. And yet, so far as I am aware, no one has hitherto attempted either to verify or to disprove the deductions by a first-hand empirical check. The material collected during our surveys of the London school population seems eminently suited for such a purpose.

At my suggestion, Miss M. Stevenson, a student in the department, made a preliminary analysis with a small, but representative sample ((17), p. 182). In the present paper, with a view to reducing the probable errors, I propose to take figures from a group that is considerably bigger than hers, though perhaps not quite so well balanced. It consists of 1784 boys, aged for the most part between $10\frac{1}{2}$ and 11 years, and attending L.C.C. elementary schools. They form part of a very large sample, of varying age, used in restandardizing a set of tests which I was originally asked to prepare for the County of Northumberland and which have since been published under the name of the 'Northumberland Group Tests'.

The marks to be analysed here fall under the usual four heads, namely, (1) general intelligence, (2) English, (3) arithmetic, and (4) handwork (including writing, drawing, and other manual subjects). Where possible, the marks for English and arithmetic are based on those obtained at the scholarship examinations. The papers set on different occasions, however, had to be equated or supplemented by

means of the standardized group tests just mentioned. Most of the children included in Miss Stevenson's sample were included in this larger group: Miss Stevenson, however, was concerned with the marks obtained in the standardized tests. I have been further indebted to her for much assistance in checking my own calculations with this larger group. Since we are here concerned only to test the results of the statistical procedure, it is unnecessary to describe in detail either the tests themselves or the manner in which the marks were awarded or rescaled: such matters have been fully discussed in previous reports.

For the primary purposes of the original inquiry, the teachers were asked to grade all these pupils in respect of their suitability (*a*) for a secondary or central school, and (*b*) for a trade school. Ordinarily such assessments are obtained only for limited groups *after* the more suitable children have entered the new type of school. But to test the mode of selection itself it seemed desirable to obtain assessments for a complete and unselected sample, including those who would be rejected as well as those who would be successful; and it was desirable to secure these assessments before any knowledge of the examination results was available, and before the pupils themselves had been dispersed to the different schools. In consequence, this larger group is somewhat younger than the group described in the earlier section of this paper.

Table 3. *General population. Observed figures*

Test	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Intelligence	0·987	—	0·751	0·593	0·505	—
English	1·403	0·751	—	0·646	0·358	—
Arithmetic	1·745	0·593	0·646	—	0·375	—
Handwork	1·502	0·505	0·358	0·375	—	—
Suitability (secondary)	1·121	0·704	0·728	0·636	0·206	—
Regressions	—	0·416	0·326	0·261	−0·220	0·806
Suitability (trade)	0·913	0·421	0·276	0·391	0·622	—
Regressions	—	0·155	−0·162	0·204	0·523	0·654

The standard deviations and correlations for the entire group are shown in Table 3. Out of the total number, 257 boys subsequently entered central or secondary schools, and another 116 were later accepted at trade schools. The data for each of these selected groups will be analysed separately; and, to begin with, the analysis will be carried out as though these were the only groups for which complete data were obtainable. We shall thus be in a position to see what deductions we should have made had we in fact been limited in our information; and later we can compare those deductions with the actual facts. In the preceding section we considered a case where *all* the inter-correlations in the general population had to be estimated. The data just described will enable us to illustrate the commoner types of case in which some or all of the inter-correlations are given at first hand, as part of the initial premises.

(*a*) *With inter-correlations not completely given*

We may take first the problem of weighting the tests with a view to determining suitability for work of a more academic type, i.e. suitability for a secondary or central school. Although in some ways it is not so closely related to our main purpose, there will be an advantage in beginning with this problem first, since the data are more reliable and the factors better known. We are to suppose that the assessments for what may be briefly termed 'secondary suitability' are available for the selected sample only. These assessments provide the criterion by which we are to judge the value of our four tests, both separately and in combination.

The boys entering the secondary and central schools were selected on the basis of the tests for English and arithmetic alone. Our task, therefore, will be, knowing (i) the inter-correlation and the standard deviations for these two subjects in both total population and selected sample, (ii) the inter-correlations and the criterion correlations for the other

tests within the selected sample only, to compute (iii) the inter-correlations, criterion correlations, and the appropriate weightings for all the tests in the total population. In this case, therefore, only one of the inter-correlations in the general population is given; and the rest have to be deduced. Having reached our estimates by pure calculation, we can then turn to the actual values (which here we happen to possess) to see how far our corrections for selection were justified.

For the selected sample (the 257 boys accepted at secondary and central schools) the inter-test correlations and the criterion correlations are shown in Table 4. In the larger or 'parent' population of 1784 children the standard deviations for English and arithmetic were 1·403 and 1·745 respectively, and the correlation between English and arithmetic, 0·646. The correction formulae are those given above in equations (vii) and (viii). The population estimates deduced with their aid are shown in Table 5. The 'true' values for the larger population, directly obtained from the test marks for the entire group of 1784, were given above in Table 3. It will be seen that, with the exception of handwork, the estimated coefficients tally quite closely with the actual.

Table 4. *Secondary school inquiry. Selected sample: data*

Test	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Intelligence	0·768	—	0·616	0·419	0·389	—
English	0·930	0·616	—	0·461	0·190	—
Arithmetic	1·178	0·419	0·461	—	0·254	—
Handwork	1·451	0·389	0·190	0·254	—	—
Suitability (secondary)	0·881	0·556	0·546	0·456	0·094	—
Regressions	—	0·372	0·242	0·228	−0·152	0·655

Table 5. *Secondary school inquiry. General population: estimated figures*

(Figures in brackets are data)

Test	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Intelligence	0·967	—	0·774	0·618	0·471	—
English	(1·403)	0·774	—	(0·646)	0·326	—
Arithmetic	(1·745)	0·618	(0·646)	—	0·384	—
Handwork	1·524	0·471	0·326	0·384	—	—
Suitability (secondary)	1·090	0·713	0·721	0·647	0·248	—
Regressions	—	0·379	0·294	0·273	−0·131	0·791

To make future selection as accurate as possible, we must determine the partial regression coefficients for the total population (or their proportional values), and use them to weight the four tests. As we have explicitly calculated all the estimated inter-correlations, it will be a simple matter to take the inverse of that matrix; then, on post-multiplying it by the estimated criterion correlations (in accordance with equation (i)), we reach the figures shown at the foot of Table 5. The values for these regressions, as computed directly from the 'true' correlations in the general population, were given in the sixth line of Table 3.

With one exception the estimates come much nearer to these first-hand values than the regressions found in the sample. The largest discrepancy occurs in the case of handwork. We may estimate the standard deviation of such a difference as being $\pm 0{\cdot}046$. It would seem, therefore, that these and the other deviations are no larger than we should expect as a result of sampling.

Our final check will consist in computing the value of the multiple correlation which we should anticipate on using the three alternative sets of regression coefficients as

weights. Had we confined ourselves to data from the secondary sample alone we should have obtained a multiple correlation of 0·655 with a standard error of ±0·036. With the weights duly adjusted for selection, the correlation rises to 0·791: the increase is fully significant. Had we taken the data actually obtained from the larger population and used the 'true' weights, we should have obtained much the same figure, namely, 0·806: the slight difference between the two correlations is not significant.

(b) With inter-correlations completely given

I now turn to the case in which all the inter-correlations for the several tests, both in the selected sample and in the general population, are given in the data. This can be illustrated by the problem of weighting the several tests to determine which pupils in the general school population are most suitable for trade schools. The boys subsequently transferred to trade schools were accepted after a varied examination intended to assess, not only their attainments in English and arithmetic, but also their general ability or intelligence and their manual and mechanical aptitude as shown by handwork, drawing, and the like. The trade school examinations, however, at any rate at the time of this inquiry, had a reliability decidedly lower than that of the junior county scholarship examinations, and far lower than that of the standardized tests. Moreover, there was here a longer interval before the actual selection was eventually made. However, for the purposes of our general argument, the effects of selection on the four main qualifications can be sufficiently assessed by taking the standard deviations and the inter-correlations for the same four tests as before, first for the entire group of 1784 and then for the smaller group ultimately selected. The figures for the selected sample, including the correlations of the four tests with the criterion (in this case suitability for a trade school) are shown in Table 6.

Table 6. *Trade school inquiry. Selected sample: data*

Test	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Intelligence	0·957	—	0·727	0·453	0·246	—
English	1·233	0·727	—	0·385	0·162	—
Arithmetic	1·426	0·453	0·385	—	0·170	—
Handwork	1·285	0·246	0·162	0·170	—	—
Suitability (trade)	0·804	0·264	0·172	0·253	0·430	—
Regressions	—	0·139	−0·046	0·144	0·379	0·478

Table 7. *Trade school inquiry. General population: estimated figuree*

	S.D.	Correlations				Multiple correlation
		Intelligence	English	Arithmetic	Handwork	
Suitability (trade)	0·856	0·406	0·307	0·369	0·529	—
Regressions	—	0·135	−0·050	0·165	0·417	0·566

In this inquiry, what we do not know, or rather are assumed not to know, are (i) the correlations of the four tests with the criterion in the general population, and (ii) the extent to which the standard deviation of the criterion has been reduced by selecting the more suitable pupils. These, therefore, are the figures we first require to estimate.

Although the number of variables is now different, so that with Pearson's formulation we should have to apply different equations, in matrix notation the formulae will be the same as before. Equations (vii) and (viii) give the criterion covariances and variance; and from these we can at once deduce the criterion correlations. These and the regressions are subjoined in Table 7.

On comparing these figures with the last two lines of Table 3, we observe that the resulting agreement is no longer so close. There are several reasons. First, the number in the selected sample is much smaller. Secondly, the assessments for trade suitability and particularly the marks for handwork are undoubtedly less reliable: in the ordinary elementary school the teachers' opportunities for assessing handwork and mechanical aptitudes are far less adequate than for assessing the more academic abilities; and, until the children are older and have had a more systematic training in manual and mechanical work, it is probable that the individual differences in this direction do not declare themselves so plainly. Finally, the actual working of the factors that have caused the selective changes may not be completely consistent with the hypotheses on which the selection formulae have been based.

The largest discrepancies arise in English and in handwork. Both affect the figure in the first decimal place; but neither is equal to twice the estimated standard error of the differences. Of the four discrepancies, two are larger, and two are smaller than their estimated probable error. As based on the estimated regressions the multiple correlation for the general population is 0·566; this is far higher than the figures obtained in the sample. As based on the empirical values for the regressions directly calculated from the observed figures it proves to be 0·655. Considering the factors just mentioned, and particularly the relatively small size of the sample, a divergence of this kind was only to be expected.

VI. PRACTICAL COROLLARIES

There are many practical implications to be drawn from the foregoing inquiries. One or two deserve explicit mention.

First, it is now evident that, by employing two different weighting patterns, we can use the same battery of tests for two very different purposes—e.g. to pick out pupils with academic ability for the secondary schools and at the same time pupils with mechanical ability for the trade schools. It follows that, for purposes of vocational guidance both in the Army and elsewhere, a proper combination of different tests may avoid the necessity of using special test batteries for different types of occupation. Here the justification for this procedure becomes clear when we make a factor analysis of all the correlations in Table 3—criterion correlations as well as inter-test. With the method of simple summation and successive approximation ((13), p. 462), we reach the factor saturations shown in Table 8. The figures imply that

Table 8. *Saturation coefficients for the general population*

	Tests					
Factor	Intelligence	English	Arithmetic	Handwork	Secondary suitability	Trade suitability
i	0·843	0·810	0·727	0·618	0·773	0·614
ii	0·114	0·370	0·131	−0·513	0·369	−0·471
iii	0·140	0·183	−0·176	0·152	−0·164	−0·135

the correlations may be accounted for by two—possibly three—main factors: first, a general factor of 'intelligence' entering into every test; secondly, a bipolar factor, classifying the children tested into those tending towards an academic type and those tending towards a mechanical type respectively; and thirdly, a minor factor possibly distinguishing verbal from arithmetical abilities—a factor of doubtful significance. Hitherto, the over-emphasis placed on the general factor (which is undoubtedly dominant) has led many to under-estimate, and indeed often to deny, the possibility of measuring more specialized aptitudes. This notion is widespread; yet it seems plainly contradicted by the foregoing results. If so, their importance for vocational testing, and particularly for psychological testing in the Army, need scarcely be pointed out. But what I want chiefly to stress is, not merely that specialized abilities as well as general intelligence may be reliably assessed, but that *several* different abilities—both 'group factors' and a 'general factor'—may be reliably assessed with the *same* battery.

The presence of negative regressions in both the investigations is a second point calling for special comment. At first sight, to propose that, in choosing pupils for schools of a higher or more specialized type, we ought actually to deduct marks if they are good at certain subjects—to handicap those who are good at English when choosing for trade schools, and those who are good at handwork when choosing for secondary—may seem almost a paradox. A moment's reflexion will make the principle clear. In the ordinary examinations for trade schools English and arithmetic are introduced, not only because certain minimum attainments in those subjects are desirable even for such practical trades as engineering, but also because the children who do best in tests of English and arithmetic are likely to be the brightest, i.e. to have a high level of intelligence. But when a reliable test of intelligence is itself included, the examination in English can add little to our knowledge of the children's intelligence. Its sole effect therefore is to introduce a new irrelevant factor, namely, specific excellence in English subjects as such; and the inevitable effect of this will be to obscure the gradings for intelligence and for manual aptitude. Indeed, the child who is *specifically* good at English (that is, better at English than his general abilities would by themselves warrant) is likely to be a bookish child with little time or interest for non-academic subjects like handwork. The specific factor of arithmetic *is* of some importance, since measurement and calculation enter into nearly every type of skilled handwork. And here it is important to emphasize one respect in which personnel selection, especially in the fighting services, differs from educational selection: in wartime training courses have to be shortened to a bare minimum; moreover, in the warfare of today the manipulation and maintenance of modern weapons and machines involves instruction in theory as well as in practice; and for this to be absorbed with sufficient speed the men must be selected for pre-existing education and experience, as well as for potential aptitude.

In the educational field, just as English and arithmetic are normally introduced in examinations for trade schools, so many have urged the introduction of handwork into scholarship examinations for secondary schools of an academic type. Here much the same arguments again hold good. As long as a direct test of intelligence is not employed, a test of handwork will help to counteract the verbal and bookish character of the examination in English and arithmetic; but when an intelligence test is added, the test of handwork not only loses its value, but actually tends to overlie the more relevant factors in the other tests.

This leads me to stress a third corollary. In earlier reports I have drawn attention to one serious defect characterizing the methods of selection commonly in vogue up to now. At the age of about 11, by means of an examination intended to test innate all-round ability, we cream off all the more intelligent children and send them to the secondary school, though many of them are probably far more suited for technical work than for academic. Then, at a later age, we re-sift the residue for pupils possessing technical interests and aptitudes; naturally we can find hardly any among them who combine such interests and aptitudes with high all-round ability, because these have already been transferred elsewhere. In a 'mechanized age' the grave disadvantages of this type of early vocational guidance are obvious.

The inference is: Either examinations for trade schools and technical institutes should be placed earlier, or examinations for scholarships and free places at secondary schools should be placed later. The psychologist, I think, must favour the second alternative. Factor analysis of test correlations year by year indicates that during the developmental period a gradual specialization of interests and aptitudes, which is scarcely discernible at all during earlier years, progressively reveals itself, and is not completed until puberty has long been passed ((5), p. 266). If these indications are confirmed, they would appear to suggest that the present age and the present procedure in the scholarship examination is apt to carry with it a premature and a one-sided specialization in the child's subsequent education, which is against the best interests both of the individual and of the community. However, these practical corollaries are merely incidental inferences. My main purpose was not to deduce educational conclusions, but to illustrate statistical methods, capable of a far wider application.

VII. Summary

1. In validating any set of psychological tests, it is important to consider both the way in which the tests are, or may be, weighted, and the way in which the persons tested have been deliberately or unavoidably selected.

2. For theoretical purposes, where high exactitude is desirable, the best method of weighting the tests is to use the partial regression coefficients, corrected, if necessary, for

the effects of selection on the sample tested. A simple formula has been given which enables the proportional weights to be calculated with little trouble.

3. For practical purposes it is often unnecessary to use expressly calculated weights, such as those based on the partial regressions: it may be sufficient merely to eliminate the unintentional weighting entailed by the differing standard deviations of the several tests, and then take a straight sum or simple unweighted average of the marks. However, where the object is to allocate persons to two or more different types of occupation, the use of two or more sets of differential weights may enable one and the same battery of tests to be used for several kinds of selection.

4. Often the groups of persons on which selection tests have to be validated have themselves been selected for those very tasks for which the tests are being devised. This kind of selection is likely to lower the correlations of the tests with each other, and to distort both the correlations with the criterion and the regression coefficients. Accordingly, it becomes important to use, not the coefficients directly obtained from the selected group, but estimated values for the unselected population. Simple formulae for calculating such estimates have been described and illustrated.

5. Since the assumptions on which the correction formulae are based are never exactly fulfilled in practice, it seemed desirable to check their applicability by empirical inquiries. It has been shown that, when the estimated values can be compared with the actual values, the results tally as closely as could be expected consistently with the probable errors of sampling.

REFERENCES

(1) PEARSON, K. (1902). "On the influence of natural selection on the variability and correlation of organs." *Phil. Trans. Roy. Soc.* A, CC, 1–66.

(2) PEARSON, K. (1914, 1931). *Tables for Statisticians and Biometricians*, Pts. I and II. London: Biometric Laboratory, University College.

(3) BURT, C. (1914). "The measurement of intelligence by the Binet tests." *Eugenics Review*, VI, 36–50, 140–52.

(4) BURT, C. (1917). *Three Memoranda on the Distribution and Relations of Educational Abilities.* London County Council: Report No. 1868.

(5) BURT, C. (1921). *Three Memoranda on Mental and Scholastic Tests.* London County Council: Report No. 2052.

(6) SOPER, H. E. (1922). *Frequency Arrays.* Cambridge University Press.

(7) KELLEY, T. L. (1923). *Statistical Method.* New York: Macmillan Co.

(8) HOLZINGER, K. J. & FREEMAN, P. N. (1925). "The interpretation of Burt's regression equation." *J. Educ. Psychol.* XV, 577–82.

(9) AITKEN, A. C. (1934). "Note on selection from a multivariate normal population." *Proc. Edinb. Math. Soc.* IV, 106–10.

(10) FISHER, R. A. (1934). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

(11) HARTOG, P. J., RHODES, E. C. & BURT, C. (1937). *The Marks of Examiners.* London: Macmillan.

(12) SNEDECOR, G. W. (1938). *Statistical Methods.* Iowa: Collegiate Press.

(13) BURT, C. (1940). *Factors of the Mind.* London: University of London Press.

(14) BURT, C. (1941–2). *Memoranda on Psychological Testing in the Army.* (Unpublished.)

(15) BURT, C. (1942). "Psychology in war" (Presidential Address, Brit. Psychol. Soc.). *Occup. Psychol.* XVI, 95–111 (abridged).

(16) BURT, C. (1942). "The value of statistics in vocational psychology." *Occup. Psychol.* XVI, 164–74; XVII, 25–33.

(17) STEVENSON, M. (1942). "Validating tests for educational and vocational selection." *Brit. J. Educ. Psychol.* XII, 182.

(18) EMMETT, W. G. (1942). *Prediction of Secondary School Success.* London: University of London Press.

(*Manuscript received* 15 *February* 1943)

TYPES OF LEARNING IN INSECTS AND OTHER ARTHROPODS

By W. H. THORPE, *From the Department of Entomology, Zoological Laboratory, Cambridge*

PART II

V. *Latent learning* (pp. 20–23). VI. *Homing and related phenomena as latent learning* (pp. 23–26). VII. *The evidence for insight learning* (pp. 26–31): (1) *Orientation and insight learning* (p. 26); (2) *The use of tools as evidence of insight learning* (p. 28); (3) *Ant behaviour and insight learning* (p. 29); (4) *Nest repair among the Eumenidae as evidence of insight learning* (p. 29).

V. Latent learning

To return to the behaviour of *Nemeritis*. It has been shown that habituation will serve to explain the production of *tolerance* to a new odour. The question now is: How far can the principle serve us in accounting for certain further results, namely, the positive response induced in *Nemeritis* to the previously neutral odour of a strange lepidopterous larva?

Nemeritis canescens (Grav.) (Hymenoptera, Ichneumonidae) is a monophagous parasitic insect normally laying its eggs only in the body of the larva of *Ephestia kühniella* (Zell) (Lepidoptera). When seeking out these larvae for oviposition the parasite is apparently guided entirely by its sense of smell. Although these insects are not naturally responsive to the odour of other related lepidopterous larvae, yet experiments using an olfactometer of the McIndoo type (Thorpe & Jones, 1937; and Thorpe, 1938) have shown that the females of *Nemeritis canescens* can be rendered positively responsive to the odour of a host larva of strange species (e.g. *Meliphora grisella*, Lepidoptera) either by rearing them artificially upon it or by exposing them, shortly after emergence as adults, to an air stream bearing the odour of the larva, without actual contact being allowed (see Part I, Fig. 1).

At first sight some might be inclined to the view that this is merely another example of habituation; the insect having become 'used to' the presence of an odour, responds by an increased rate of change of direction on its removal and so appears in the olfactometer to 'choose' the odour. But a moment's reflexion will show that such an explanation has an apparent simplicity which is in fact specious. The theory in reality involves an extension of the concept of habituation to cover complex phenomena of a very different nature and thus entirely robs the term of its original meaning, indeed of any precise meaning. The essence of habituation is the waning of an inherited or instinctive response. In this experiment, however, we start out with no response, *Nemeritis* being totally indifferent to the odour of *Meliphora*. What requires explanation here is the building up, by some process other than simple association, of a new response from zero, not the waning of a pre-existing response, and clearly habituation cannot apply in such a case. Nor does it appear possible to hold a mixed odour hypothesis, in the sense outlined above (see Part I of this paper) for the *Drosophila*-peppermint experiments, since it appears extremely improbable that *Meliphora* larvae would give off an odour that is actually repellent or irritant. Still more improbable is it that the attractant and irritant odour should be so

exactly balanced as to give consistently neutral effects over a very large number of experiments (see Thorpe & Jones, 1937, Tables III, IV and XII; and Thorpe, 1938, Tables I, VIII*d* and XII). How then can this result be accounted for? With what other examples of animal behaviour is it comparable? A study of the contemporary literature of invertebrate behaviour provides little that illuminates the problem. But when we consider the great mass of work on maze running with rats and other mammals, our attention is at once arrested by the phenomenon known as 'latent learning'. The first clear evidence that in rats learning can take place in the absence of a reward was provided by some experiments of Lashley (1918), who showed that when a rat is given food in a somewhat unfamiliar environment (e.g. the food box of a maze) it will almost invariably explore the neighbourhood of the food before eating. In the circular maze the exploration usually includes the food compartment and the alley surrounding it. Once the exploration is completed the animal will go directly to food in succeeding trials. Much more striking evidence was provided by the experiments of Blodgett (1929), Elliott (1930) and Tolman & Honzik (1930) who showed that rats which had had daily unrewarded runs through a maze on ten successive days, when presented with food in the food box displayed a very striking decrease both in errors and times of running as compared with a control group (see also Haney, 1931). Maier (1932) showed that a very simple maze may be entirely mastered through random exploration; the introduction of a reward therefore merely causes the animal to demonstrate its learning. Further confirmation of this was provided by Daub (1933) and Herb (1940). The fully 'trained' animal, once it has found food in one part of the maze can then go direct to that point from any other part of the maze in which it is placed.

It is suggestive that a conclusion somewhat analogous to the outcome of the latent learning experiments is arrived at as a result of the illuminating work of Brogden (1939) on 'Sensory Pre-Conditioning', working with dogs and using the classical conditioned reflex technique of Pavlov. This worker showed that if dogs be given successive stimulations of sound and light in combination contiguously, and if either sound or light be then made a conditioned signal for forelimb flexion from an electrically charged grid, the stimulus not involved in conditioning will then invoke flexion without being combined with shock. But how far is latent learning entitled to the distinction of a separate category and what are its essentials? And if it is a distinct type how is it to be related to the other categories of learning? That it represents something more than association has been demonstrated by Koffka (1935, pp. 588–9). Maier & Schneirla (1935, pp. 404–6) apparently take the view that latent learning is evidence of a simultaneous integration and reorganization of previous experience and that therefore it must be considered as essentially insight learning. That it is evidence that the maze pattern must, by some process, eventually be learned as a whole seems highly probable; but this does not *necessarily* involve insight learning. I do not think therefore we can consider latent learning as equivalent to insight learning proper since the addition of an incentive in the form of food in the food box does not make the situation essentially a new one. If latent learning be accepted as evidence of insight learning, then it is difficult to deny insight learning to many animals capable of homing. All that in fact happens is that certain parts of the maze are rendered relevant to the problem of satisfying hunger; so that the situation is only 'new' in that the animal's *attention* is now concentrated on certain parts only of its previous environment. Similarly, the insect in the olfactometer is in a new situation only in the sense

that it is in a simpler, more limited environment—one where the odour is localized and so capable of being responded to—and immediately, without any trial and error procedure, the *Nemeritis* is able to act upon its previous experience, tending to use the odour as a signal for a total environment which contains substances that can satisfy certain of its needs. Prior to its training its sense organs could register the presence or absence of the odour but the odour had no significance in its life and it did not attend to it. Now the odour has acquired a 'meaning' and the insect exhibits the phenomenon we call '*attention*'.[1]

It seems then that latent learning involves the following four separate elements, the last three of which are characteristic, in that they are not necessarily found in the simpler types of learning already dealt with. Firstly, as mentioned above, latent learning involves integration of a pattern as a whole; secondly, it is learning without specific motivation by any of the primary needs; thirdly, it results in a kind of 'transfer of training' whereby the animal can use its learned behaviour in response to changed motivation; and fourthly, it connotes a faculty enabling the animal, without any trial and error process, to 'select' or attend to certain parts of the previously learned whole as a signal for the whole or in relation to certain specific needs.[2] Some might be inclined to call this last process 'abstraction', but since it is not equivalent to and does not necessarily involve *insight learning*, the term 'abstraction' seems best avoided.

With regard to the question of motivation it might be objected that there can be no action or learning without it and that the animal, if neither seeking food nor motivated by one of the other primary needs, is trying to get out; in other words, is activated by the 'need' to escape. To use Russell's (1935, 1938) terminology, the whole environment has escape valence. But the experiments do not in fact support this conclusion, for the rat by its efforts does not get food nor does it escape (see Haney, 1931; Daub, 1933; Herb, 1940); therefore on the trial and error principle there is no basis for learning. Actually what we have here amounts to much the same thing as 'exploration for its own sake'. We may say that 'the animal tends to explore its surroundings'; that it has a tendency to investigate its environment. It is this phenomenon to which Pavlov refers in his singularly inappropriate term 'investigatory reflex'; inappropriate, in that if this complex behaviour is described as a 'reflex', the term 'reflex' loses all precise meaning and becomes a synonym for response. 'Latent learning' may also be an unfortunate term but we shall make a great mistake if having dubbed the phenomenon 'exploration' or 'investigation' we leave it at that and assume that all is fully understood and the observations of little significance. On the contrary they are of the greatest significance; for in describing a piece of behaviour as 'exploratory' or in speaking of 'an exploratory instinct' we are at once placing the organism on a relatively high behavioural level.

Indeed we must expect to find that animals which are capable of any *complex* homing behaviour are, *mutatis mutandis*, psychologically in advance of those which are not. Thus the behaviour of the rat displaying latent learning, the behaviour of *Nemeritis* in the olfactometer, and those more complex types of homing orientation exhibited by certain bees and wasps, where we have definite orientation flights (to be dealt with in the next section), appear to be, in essentials, identical. If an insect can build up the discrete

[1] This learned or acquired attention is of course something over and above the so-called 'attention' based upon the interference of unlike reflexes and the alliance of like reflexes in their action upon the common paths (Sherrington, 1911).

[2] Compare the phenomenon formerly termed 'apperception' by psychologists.

elements of its perceptual world into a single pattern which thenceforth signifies a locality which is, in some sense of the word, a 'home'; and if, as appears in certain instances to be the case, it can then, as need arises on its foraging flights, use now one, now another part or aspect of this perceived pattern for identifying its base of operations: then we have behaviour which can tentatively be equated with 'latent learning'.

VI. HOMING AND RELATED PHENOMENA AS LATENT LEARNING

As is well known, many invertebrate animals, including several diverse genera of molluscs and possibly even echinoderms (Russell, 1938), exhibit traces of an ability to return to a particular locality which in some senses may be described as 'home'; in insects usually, but not always, associated with the care of the offspring. In some cases it is possible that the homing faculty depends on no more elaborate sensory mechanism than that involved in simple taxes or the light compass reaction. But in many other cases a true 'place memory' seems to be involved. Thus in the mollusc *Patella*, Piéron (1909) has shown that the combination of at least three different types of sensory impression is involved (see also Arey & Crozier, 1918, on *Onchidium*). Similarly, the more complex type of homing orientation such as that found in the Heliconine Lepidoptera (Jones, 1930, see p. 24 below) and in the social Hymenoptera and the Sphecoidea, while clearly embodying simple reactions, appears in many cases to be effected (once the general neighbourhood of the 'home' has been reached) by a co-ordination of a variety of sensory impressions (Rau, P. 1929; von Frisch, 1931). This type of complex orientation is of course a very similar activity to that of learning to run a maze, and in many instances we cannot rule out the possibility that the path to and from the nest or burrow is learned piece by piece just as a maze may conceivably be learned by a rat, by building up a series of responses starting with the food box and its immediate surroundings and working backwards therefrom. But I think it can be shown that this can hardly be all, and that in certain instances the integration of the environmental pattern into a single whole which can then be used in a variety of new ways must be postulated.

So much has been written on the orientation of the Hymenoptera, and summaries are so easily accessible in text-books such as those of Bouvier, of Schröder, of Rabaud, of Warden, Jenkins & Warner and of Fraenkel & Gunn (see also Turner, 1923), that we need hardly take space in describing in detail numerous examples of the phenomenon. It seems generally established that under natural conditions the return journey to the nest falls into three fairly well-demarcated sections. Firstly, there is the departure from the food place in the correct direction; secondly, the location of the nest site or environment of the nest as a whole; and thirdly, the recognition of the nest itself. As many observers have pointed out, the first part of the journey may be very different from the rest in that at the start the nest is often out of sight and in fact cannot be perceived at all. It now seems fairly clear that the taking up of the correct direction at the commencement of the return is not due to any mysterious 'direction sense' but is, as Cornetz (1911) says, a function of the going, and depends in some way on the internal or external stimulation involved on the outward journey. The light compass reaction or more complex visual cues are thus in certain instances essential for the right start of the return trip just as they are for its completion, even though the goal may be well out of sight; and kinaesthetic sense also undoubtedly plays a large part.

As an example of the flight orientation of the Hymenoptera let us take the nuptial

flight of the queen honey bee, where one or two short 'orientation flights' prior to departure are all that is needed in preparation. Such a performance considered in terms of maze learning is seen to be an astonishing achievement. It seems that the bee must respond to the 'home' area as an entity apprehended as a whole, utilizing certain landmarks as orientation signals for it.[1] There are many studies on the orientation of the Hymenoptera which support this conclusion. Thus the observations of Bouvier (1901), Buttel-Reepen (1907), Friedlander (1931), Rabaud (1924, 1926), Turner (1908) and Wolf (1925, 1930) all emphasize the importance of the general pattern of the site as a whole.

Wolf describes how a worker bee taken from the hive and transferred in a box to any point within the normal foraging area is able to recognize its location and find its way back to the hive without difficulty.[2] The importance to foraging bees of the general pattern of the site is also clear during that very critical period of training experiments when the workers are first being taught to come to a feeding table. At this time the bees are apt to be confused by very slight changes in the surroundings of the table (M. Hertz, personal communication), and at this stage the experiments are very easily brought to nought. Later, when the routine is well established, kinaesthetic factors come to be more and more relied upon. Similarly, with *homing* bees and wasps when the nest has been moved, returning workers may, as is well known, hover where it was formerly situated although after a while they may recognize it at another point. Wolf has shown, with hive bees, that it is the older workers with more stereotyped kinaesthetic habits, which tend to behave thus. Younger bees, on the other hand, and those deprived of the antennal organs of kinaesthetic sense, show greater adaptability and more easily find the hive in its new situation. But because kinaesthetic factors are important we must not therefore assume (as some have done) that the insect is no longer orientating itself by means of landmarks and the general pattern of the site. On the contrary, the kinaesthetic sense is the means by which the bee uses its landmarks with the greatest economy of effort. Insects are in general very short sighted, and the kinaesthetic sense enables the bee to bridge greater gaps from one landmark, or constellation of landmarks, to the next than would be possible by vision alone. The work of Brun (1914), Descy (1922), Ferton (1901–21), Hartman (1905), March (1930), Opfinger (1931), the Peckhams (1900, 1905, etc.), Rabaud (1919), Tinbergen (1932, 1934) and Verlaine (1924–34) makes clear the importance of landmarks in the recognition of the site. In fact homing insects may react either to one or more of the various aspects or features characteristic of a situation or to a complex combination thereof, i.e. to the situation as a whole. Thus many instances of flight orientation in the Hymenoptera, particularly those where there are distinct preliminary 'orientation flights', appear to be psychologically equivalent to the process of latent learning as manifested by rats in a maze. The work of Opfinger (1931) on the honey bee merits particular attention in that it makes clear the distinction between the technique of arrival and of the outward flight, the food being on a yellow ground upon arrival and upon a blue ground upon departure. If later the two colours are presented beside each other the insects land only on the yellow ground and are confused if the blue is presented alone.

A very striking evidence of the 'place memory' of gregarious butterflies, *Heliconius charithonia*, which home to their roosting place as an entity apparently apprehended as a

[1] Just as does the pigeon to its loft (Watson & Lashley, 1915), or the wild bird to its territory (Howard, 1935, p. 78; 1940, p. 11).

[2] See also Molitor (1937*b*, 1939) on homing experiments with Vespidae and Sphegidae.

whole, is supplied by Jones (1930, 1931). Other examples of gregarious sleeping habits of insects (see Rau, P. & N. 1916) may well be found to involve orientation responses of a similar type.

The ability of some wasps and bees (Tinbergen, 1934; Molitor, 1939, etc.) eventually to recognize a nest after it has been removed to a new site has been mentioned above. Now since the insect is at first homing to the whole local environment plus the nest this well-known fact has seemed to some workers to indicate a considerably greater power of selection than is involved in the account of orientation just given. Verlaine (1932–4) claims that wasps and bees can be taught to identify food and nest places by means of response to simple geometrical figures, and indeed that bees, in distinguishing the triangle from the square and the circle, show a power of abstraction such that a generalized feeding reaction to thirty-four kinds of triangles can be obtained! This result is, however, contrary to the results of Hertz (1929, 1937) and of many others who have worked on the form sense of bees. In fact, it appears that while bees can distinguish complex patterns they fail to discriminate between simpler geometrical forms which they are unlikely to meet in their normal environment. In view of this conflict of results and in view also of other results of Verlaine's (see Lashley, 1940), his conclusions should be regarded as requiring confirmation.

It seems that the general process of homing orientation in ants is, allowing for the great differences in sensory equipment, similar to that in bees and wasps; and our discussion of ant orientation under the experimental conditions of maze running (Part I, above) has in fact covered a large part of the field. For further references to work on both groups the reader is referred to the text-books cited above, the comprehensive work of Warden, Jenkins & Warner (1940) having a particularly extensive bibliography. The orientation behaviour of the termites seems to be similar in essentials to that of the more primitive species of ants (Goetsch, 1936), but comparatively little has yet been done.

To return to the experiments with *Nemeritis*. Obviously the acquired positive response to a new odour must be a generalized reaction, that is, an attempted return to a favourable environment as a whole, for which odour has now become a signal. It cannot be a specific need response, linked directly with the process of satisfying one of the primary needs (e.g. food, reproduction, etc.), since these possibilities were designedly eliminated by the conditions of the experiment. The case is therefore in many respects similar to a homing reaction, differing only from the usually accepted type of homing in that the experience leads the animal back, not necessarily to an identical spot but to a particular type of environment.

That this kind of 'homing' to a type of environment rather than an exact locality, as induced experimentally in *Nemeritis*, may be a widespread and important phenomenon of learning among insects has already been discussed elsewhere (Thorpe, 1938, 1940). Thus there may well be numerous biological races, based on conditioning to particular food plants, which are entirely phenotypic in character. Since such food plant conditioning will tend to isolate those members of the species which possess it from other members associated with other food plants, it will *ipso facto* tend to bring about a non-genetic mating restriction or preference—a factor which may prove to be of very great evolutionary significance. These facts, and the recent work on bird behaviour (see Cushing, 1941*a*), give a new colour to the idea inherent in the work of earlier writers (e.g. Gulick, 1905), that acquired non-heritable behaviour may have great influence on the course of evolution.

But obviously only careful investigation can enable us to determine for certain whether a given instance of seeking a special environment is phenotypic; and in very few cases has the necessary information been obtained. Thus the attraction of parasitic Hymenoptera to the food plants of their hosts (Thorpe & Caudle, 1938) may prove to be either instinctive or acquired or something of both.

The only other clear instances of homing behaviour in insects which need mention are those of the Odonata and certain aquatic Hemiptera and Coleoptera. Though no careful observations on this subject appear to have been made, every collector of dragon-flies is familiar with the fact that certain species show marked territorial behaviour when hawking for their prey and will haunt the same circumscribed terrain for days at a stretch. Here no doubt the orientation is on a visual basis.

The behaviour of some surface-living aquatic insects comes very near to that of the Odonata. Thus gyrinid beetles, even though describing their whirling courses apparently at random on the surface of flowing water, maintain a constant general position by means it seems of visual perception of the pattern of objects on the adjacent banks. Since the pattern of visual stimulation is constantly changing with the movements of the beetle, this seems to be essentially a homing reaction; and it is known that the same flotilla may occupy an identical location for the whole of the season or even from one year to the next. How far the orientation is to a pattern of objects apprehended as a whole is, however, not quite clear. Brown & Hatch (1929) state that the beetles are completely disorientated if a sudden change in landmarks is made. A simpler instance is provided by *Notonecta* (Schulz, 1931), which in a flowing stream can keep its head to the current and remain stationary by swimming in such a manner that its visual impressions are kept constant. In so far as such behaviour is merely the compensation for change in stimulation of the ommatidia, it is correctly regarded as an example of light compass orientation (menotaxis). However, there seems little doubt that in some cases form sight is involved.

To sum up: Observations and experiments on the homing of animals in the field cannot of course be expected to yield readily the precise evidence of latent learning which, when applied to suitable cases, the maze and olfactometer techniques can provide. But where we get, as in some of the cases outlined above, a homing response to a complex environmental pattern very rapidly learnt by the individual as a result of specific orientation excursions—yet at the same time a response which is to some extent flexible and adaptable as the situation requires—we have reasonable grounds for the tentative assumption that powers essentially equivalent to those concerned in latent learning must be involved. But such behaviour does not necessarily imply insight learning.

VII. The evidence for insight learning

The object of the present section is to consider critically those few cases of insect behaviour which appear to provide *prima facie* evidence for insight learning as defined in the introductory section.

(1) *Orientation and insight learning*

There are a few well-known cases of orientation, which might have been dealt with under 'homing', but which do seem at first sight to suggest something more than powers of the latent learning type, and which are therefore dealt with separately. I refer to the

behaviour of the sphegid wasp *Ammophila urnaria*, returning to its burrow with prey, first described by the Peckhams and later corroborated in other species by other workers (e.g. Rau, P. & N. 1918); to the *Sphex* watched by Turner (1912); and to some pompilids (e.g. *Cryptocheilus rubellus*, described by Hingston, 1928) which capture spiders.

The wasps of the genus *Ammophila* stock their nests with caterpillars. In the case of *urnaria* and some others only one or two larvae are procured for each burrow, and these perforce must be large ones. If they are not too big they are carried back in flight, but they often exceed the size at which this is feasible and accordingly they have to be dragged along the ground, over or through all kinds of obstacles, a distance which may exceed 250 ft. in some cases. From time to time during this laborious journey the wasp leaves her prey to take short survey flights, thus maintaining the correct direction. The remarkable thing here is of course the fact that the wasp has never traversed the route on the ground before and can apparently only maintain it by frequent observation from the air. Now if this orientation is entirely by means of a series of landmarks observed from the air but followed from the ground, then I think we should have to postulate a recombination of previous experience to meet a new situation. There is, however, the probability that the light compass orientation is solely responsible for maintaining direction while on the ground, the air surveys being for the purpose of relating the direction of the sun's rays to the main landmarks. Molitor (1937 *a*), however, concludes, on the analogy of Brun's (1914, 1922) experiments with ants, and from his own observations, that light compass orientation plays little or no part in the direction finding of sphegids. But, on the other hand, we know from the work of Wolf (1927) that it can operate with bees when landmarks are few. The recorded observations on the direction finding of *Ammophila* do not then provide the evidence for us to decide finally between these two alternatives, so we must for the time being accept the latter as being the simpler and more consonant with the method of orientation in other Hymenoptera.

But there is one isolated observation by Turner (1912) which stands out from the rest by virtue of the circumstances which are carefully recorded and which seem to render improbable at any rate the more obvious alternative explanations. Turner describes how an unidentified *Sphex* in dragging a spider directly towards its nest in a north-westerly direction encountered a north-south 4 ft. fence which had no gaps and which owing to a projecting scantling constituted an insuperable obstacle. A short orientation flight was made as a result of which the insect set off along the fence with its burden for 6 ft. due south to a point where its end met a much lower fence running east. This fence had gaps in it. The *Sphex* turned east along this fence for a short distance until it reached the first gap, through which it passed. It then turned due west, rounded the corner made by the two fences, and after a further short orientation flight now dragged its prey, once again approximately in a north-westerly direction, direct to its goal. Readers of the first chapters of Köhler's book (1925) will at once see that this detour behaviour suggests a considerably greater insight learning ability than that shown by the average hen, being in fact more nearly on a par with that recorded for dogs. Unfortunately, we know nothing of the previous history of the individual insect, nor was the matter followed up by further experiments.[1]

[1] Hingston (1928, p. 140) states that *Sphex* can make detours on the ground as the result of its orientation flights, but he gives no details. Nor does he, as far as I can find, mention the matter in his paper on *S. lobatus* (1925).

So far then our survey of maze learning and homing behaviour in arthropods has supplied us with only two cases giving even *prima facie* evidence for insight. Where else might we expect to find it? It will be remembered that the classic example of insight behaviour in animals is that of Köhler's apes. Are there any insect activities which show anything even remotely resembling the abilities of the higher mammals in spontaneous tool using?—for this is what much of the behaviour of the apes in these experiments amounts to.

(2) *The use of tools as evidence of insight learning*

There are certain cases of tool using among insects which have in the past been regarded as evidence of 'intelligence' and these must now be considered. Firstly, there is the famous instance of the tropical ants of the genera *Polyrachis* (Jacobson, 1905, 1908) and *Oecophylla* (Ridley, Bugnion, Dodd) which make nests and shelters of leaves spun together by means of silken threads. This silk is secreted only by the larvae which are held in the jaws of the workers and used as tools for this purpose. Now if the older psychologists were right in regarding tool using as clear evidence of the highest form of intelligence, no more need be said; the case for insects would be proved. But I think we need not assume any insight learning to be involved in the present behaviour of *Oecophylla*, difficult though it may seem to dispense with it in accounting for the *origin* of the habit.[1] The whole series of actions on the contrary is undoubtedly a piece of instinctive behaviour of a very high degree of fixity. So much so, that although the larvae pupate naked, spinning no cocoon, yet their silk glands actually show a structural modification, being markedly hypertrophied (Ridley, Doflein).

By far the most famous instance of tool using among insects is that of the sphegid wasp *Ammophila urnaria* which fills in the mouth of the completed and provisioned burrow with earth, carefully pressed down. The Peckhams observed an individual wasp which, contrary to what appears to be the normal behaviour of the species, picked up a small stone in its jaws and proceeded to use this as a hammer, ramming down the soil with swift repeated strokes till the entrance to the burrow was no longer distinguishable from the surrounding earth. A very similar observation was made earlier by Williston (1892) on *A. yarrowi*, on another unidentified species by Hungerford and Williams and by P. & N. Rau (1918) on *Sphex* (*Ammophila*) *pictipennis*; in this latter species it is apparently the normal method. Hartman (1905) records that when *Ammophila procera* is closing its nest it sometimes happens "that a piece of wood is pressed down tightly, then pulled out and pressed down again, and this is repeated several times so that one might suspect that the wasp were here improvising a tool with which to tamp down the sand". Armbruster (1921) and Minckiewicz (1933) both record a similar habit for the European *A. sabulosa*, and according to Berland (1935) all individuals of *A. heydeni* use a stone exactly after the manner of *A. urnaria*. Molitor (1931, 1937*a*), however, disagrees; he says that *A. heydeni* uses a stone "bisweilen, nicht immer".

From the original observations of the Peckhams (1898, 1900), many biologists and psychologists (e.g. Morgan, 1900; Bouvier, 1922; and McDougall, 1936) did not hesitate to describe the exceptional individual as a genius leading its species onward to the use of tools. But the subsequent observations quoted above have rather altered this picture. It is now clear that the act is one which borders very closely on what we can only regard

[1] The *origin* of instincts is itself of course one of the major problems of biology. To attempt to deal with it in this paper is obviously impossible.

as the normal instinctive behaviour of the insect; in fact in some species it *is* the normal instinctive behaviour. In all species the seizure of pebbles, clods of earth, bits of wood, etc. in the mandibles and the packing of dirt into the hole are part of the instinctive act of filling and closing. Those species and individuals which use a stone as a hammer are therefore merely combining two features of the instinctive hole-filling process in a rather unusual way. Since the animal is constantly handling and manipulating clods and stones it has every opportunity to learn by the chances of its normal experience the results of handling stones and other hard objects of various kinds. The behaviour of the Ammophilas and other sphegids and pompilids does I think unquestionably give very strong support indeed to the views of the hormic or purposive school of comparative psychologists (McDougall, 1936; Tolman, 1932). I do not see how we can possibly give careful and unbiased consideration to all the remarkable details of this complex yet flexible instinctive behaviour of the Ammophilas as set forth by P. & N. Rau and many others, without conceding conation and some element of purpose and anticipation, however slight. But that they *necessarily* involve insight learning as here defined I cannot regard as proven. For a full discussion of the significance of tool using in animals the reader is referred to the work of Ettlinger (1924).

(3) *Ant behaviour and insight learning*

It is of course a commonplace that anyone who studies the behaviour of ants will see and read of much which superficially suggests insight: for instance, the making of living bridges and the foraging activities of the army ants (Dorylinae). Yet it is a fact that, compared with those of many other groups of ants, the individuals of the doryline colonies are psychologically of a very low grade; though as a result of their instinctive equipment, highly co-operative socially. The way in which the seemingly marvellous powers of concerted action in overcoming obstacles can be described without recourse to the conception of insight learning has been demonstrated by Schneirla (1933*b*, 1934*a*, 1938, 1940) and by Lutz (1929), and their conclusions seem applicable with little difficulty to many other examples of elaborate ant behaviour (e.g. Turner, 1907*b*). Thus it is shown that (Schneirla, 1938) the alternation of nomadic and sedentary behaviour in *Eciton hamatum* depends on the state of the brood and thus ultimately upon the physiological condition of the queen. The raiding of *E. burchelli* (Schneirla, 1940) is initiated through photokinesis, and its form and type are governed primarily by the inherited behaviour pattern (the 'rebound pattern') of the individual. Individual learning (habituation and simple association) makes a minor though essential contribution. The concepts of 'pressure' and 'drainage' coupled with the transmission, based on topochemical sensitivity, of a generalized excitement through the colony (see also Eidmann, 1927 and Goetsch, 1934), account for a great deal, and it is shown that there are no individuals specialized or trained for leadership. In the advancing swarm the 'lead' arises through the function of the entire group, all individuals making essentially similar contributions to 'leadership' under equivalent conditions.

(4) *Nest repair among the Eumenidae as evidence of insight learning*

There are on record a number of experiments in which Eumenidae and Megachilidae have been induced to break off the routine of nest making and to repair artificial injuries to their nests. The authors of these experiments, Hingston (1927 *a*, *b*) and K. D. & W. McDougall (1931), the latter a psychologist of very great experience and reputation, have

regarded their observations as undoubtedly indicating insight learning. Hingston's observations are also of the very greatest interest but his use of psychological terms is so loose as to invalidate his discussion. This defect unfortunately detracts particularly from his book (1928) in which nevertheless he has performed a valuable service in gathering together a very large number of interesting cases of adaptive behaviour among insects, mainly drawn from his own experience and most vividly described.

Two of Hingston's experiments are of particular interest. From among all the examples collected in his work these are the only ones which I feel after careful examination give a *prima facie* case for postulating insight learning. They are on the lines of Fabre's experiments with *Chalicodoma*, but very different results were obtained.

Eumenes conica is an Indian 'Potter Wasp' or 'Mason Wasp' which builds a cluster of clay cells. After depositing an egg in each she fills it with caterpillars and seals it with a lid. Eventually the whole cluster is covered with a layer of clay. While the wasp is away hunting for its prey the experimenter makes a large rent in the side of the cell. The wasp returns and puts in a caterpillar which falls out through the hole. A second is brought; this one sticks in the hole, a large part of it hanging out but blocking the aperture sufficiently to prevent subsequent ones from dropping through. All this time the wasp shows no sign of having noticed the breach, her attention being absorbed in provisioning. But when all the caterpillars are in, she appears to notice the hole for the first time and examines it and the dangling caterpillar carefully. Then with great and prolonged effort she manages to stuff the larvae back; she then goes away and collects a pellet of clay with which she mends the hole.

The second instance concerns *Rhynchium nitidulum*. This wasp is a builder of pots, usually on the ceilings of rooms. Her routine is not unlike that of *Eumenes*. First she builds a clay pot, then covers it with resin, then stuffs it with caterpillars, finally sealing it with a lid. When one pot is finished, another is put alongside resulting in a cluster of pots. Hingston makes a hole in the bottom or side of a finished pot which already contains an egg and the wall of which has been smeared with resin. The returning wasp sees the cavity and makes a careful examination of it. She then makes a long but ineffectual attempt to mend the hole with bits of resin pulled from the wall of the pot. After two hours of effort night supervenes and she spends the hours of darkness in the pot. At dawn she leaves it, and on her first return visit carefully investigates the hole again. She goes off a second time and brings back a pellet of mud. Entering the cell through the hole she sets about repairing the damage. First she works at the edge of the hole, filling in all the little irregularities until the ragged opening is converted into a uniform circle. Then she closes the hole with a lamina of clay which she manages to curve neatly into the natural rotundity of the pot. She next fetches resin which she smears all over the patch until it blends with the rest of the pot. Hingston points out that the wasp repairs the hole by a method different from that employed in closing the mouth of the pot, for she does the repair by a different set of actions. "When closing the gate she sits outside and from there introduces clay. When repairing the hole, she examined it from both sides, and then, having made a choice, elected to do the repair from within." Hingston adds: "The wasp breaks from the bondage of her instinct, attacks the difficult problem intelligently, and repairs the damage with complete success."

The example recorded by K. D. & W. McDougall (1931) is so similar to that of Hingston's *Rhynchium* that a detailed account is hardly necessary. The McDougalls' account

deals with an unidentified 'mason wasp' said to be common in North Carolina. In this case too a hole was made in the bottom or side of the pot and the provisions, namely, spiders, removed. Here again a repair is carried out, the authors stressing the fact that the exact method and movements of repair are different from those observed during any part of the normal routine of pot-building and provisioning. In this case the repair is performed from the outside whereas normal construction proceeds from inside. When the repair has been completed there is a good deal of rather aimless patching up with pellets of clay after which the cell is restocked with spiders and another egg is laid.

Now I think everyone will agree that such observations do at first sight strongly suggest insight learning. The animal *appears* to give a new and adaptive response to a new situation without apparently any process of trial and error. But before we can regard such cases as positive proof of insight learning we must know for certain: (1) that the experimental situation is not one that the insect can ever meet as part of its normal experience; (2) that the behaviour that results is really new and not a piece of instinctive behaviour transferred from the behaviour pattern appropriate to some normal situation. For it is well known to all students of animal behaviour that animals under sudden emotional tension, and particularly when frustrated in some way, will often respond by some 'stupid' act, that is, some totally irrelevant piece of instinctive behaviour arbitrarily transposed from some quite different environmental stimulus pattern (see Armstrong, 1942). It is obvious, then, that to be quite certain on these two points we have to know a very great deal about the normal environment and behaviour of the animal under observation, and I do not feel sure that in any of these three cases we really have this knowledge. For instance, with regard to the making of a hole in the pot, are we quite sure that there is no bird or other predatory animal which breaks the cells open in this way and extracts the contents and to which the insect may be adapted? Hingston, it is true, states categorically, "A new experience is put before the wasp, something she has never met with before." But can he be quite certain? We know a good deal; but do we yet know enough for these experiments to be regarded as final proof of such a far-reaching conclusion?

With regard to all the other cases described in Hingston's absorbing book, some of which appear to convince him of the existence of a very high degree of intelligence, the reader must be left to judge for himself; there is no space to consider them all here one by one. But I feel convinced that any impartial critic, fully alive to the difficulties and pitfalls of such investigations, will in each case be able to provide some simpler alternative hypothesis which is not inconsistent with the facts as given by the author.

ERRATUM

In Part I of this paper (this *Journal*, xxxiii, Part 4), on p. 224, line 34, 'visualisation' *should read* 'visual generalization'.

SUSCEPTIBILITY TO A VISUAL ILLUSION, AS RELATED TO PRIMARY AND SECONDARY SUGGESTIBILITY AND OTHER FUNCTIONS[1]

By H. J. EYSENCK

I. *Introduction* (p. 32). II. *Experimental method* (p. 32). III. *Results* (pp. 33–35): (1) *Factorial analysis of Muller-Lyer illusion* (p. 33); (2) *Correlation between susceptibility to Muller-Lyer illusion and tests of suggestibility* (p. 34); (3) *Correlation of Muller-Lyer illusion with other data* (p. 35). IV. *Summary and conclusions* (p. 35). *References* (p. 36).

I. Introduction

Visual, auditory, and other illusions have often been used as tests of suggestibility. Gilbert(1) and Seashore(2) were probably the first to use illusions in this context, and Brown included several illusions among his numerous tests of suggestibility(3). Hollingworth put forward the orthodox view when maintaining that "any perceptual illusion may be said to constitute a case of suggestion"(4). Little is known, however, with regard to the correlation of 'susceptibility to an illusion' and 'suggestibility' as measured by any of the various tests of personal and impersonal suggestion in common use. The present paper reports on the relation between susceptibility to the Muller-Lyer illusion, and eight tests of suggestibility, as well as on a few data gathered from other tests.

II. Experimental method

In the Muller-Lyer illusion, the observer is required to equate two lines, one of which appears shorter because of two arrowheads affixed to the ends, while the other appears longer because of two inverted arrowheads going out from its ends. In this particular experiment, the 'arrowheads' part of the figure was held constant, while the 'inverted arrowheads' part of the figure was adjustable. The constant part was 7·7 cm. long; the oblique lines were one-quarter as long as the horizontal segments, forming an angle of 45° with them. The method of equivalents was used, the experimenter varying the movable part of the figure from a position where it was obviously longer than the constant part to a position where the observer called out 'Equal', and from a position where the movable part was obviously shorter until the observer called out 'Equal'. Each of these two ways of testing the susceptibility of the subjects was repeated three times. The three trials where the variable line was *shortened* ('IN' trials) will be denoted A_1, A_2, and A_3; the three trials where the variable was *lengthened* ('OUT' trials) will be denoted B_1, B_2, and B_3.

Sixty patients at Mill Hill Emergency Hospital were tested in this manner. Fifteen of these patients were male hysterics, fifteen were female hysterics; fifteen were male non-hysterics, fifteen were female non-hysterics. The hysterics without exception had conversion symptoms.

[1] I wish to express my thanks to Dr W. S. Maclay for his permission to use the clinical material at the Mill Hill Emergency Hospital. I am also indebted to the Rockefeller Foundation for a grant which enabled me to carry out this research.

III. RESULTS

(1) *Factorial analysis of Muller-Lyer illusion*

The six trials of these sixty patients were intercorrelated, and the resulting table factor analysed. The results are given in Table 1. Two factors are sufficient to account for all the significant correlations. (According to Fisher's method of calculating the significance of small samples, a correlation of 0·25 is significant ($p = 0{\cdot}05$), while a correlation of 0·32 is very significant ($p = 0{\cdot}01$), when $n = 60$.) The first factor is positive throughout, and accounts for 59 % of the variance, while the second factor is bipolar, and accounts for 13·5 % of the variance.

By means of Burt's adaptation of the Spearman-Brown prophecy formula, shown by the writer elsewhere to be applicable to data of this kind(5, 6), we can predict how closely the average susceptibility of our subjects on the six trials would correlate with their 'true' score, i.e. with the susceptibility score they would achieve if the experiment were repeated a very large number of times. This correlation is 0·99; in other words, six trials are sufficient to measure the phenomenon under consideration with considerable reliability.

Table 1

Trial	A_1	A_2	A_3	B_1	B_2	B_3	Factor I	Factor II
A_1	—	0·792	0·670	0·391	0·481	0·546	0·750	0·367
A_2		—	0·718	0·491	0·532	0·641	0·846	0·343
A_3			—	0·330	0·349	0·527	0·660	0·369
B_1				—	0·768	0·774	0·710	−0·460
B_2					—	0·756	0·751	−0·374
B_3						—	0·870	−0·267
						Variance	0·590	0·135

Table 2

$A_1 = 1{\cdot}8$ cm. $\pm 0{\cdot}602$	$B_1 = 2{\cdot}5$ cm. $\pm 0{\cdot}530$
$A_2 = 1{\cdot}7$ cm. $\pm 0{\cdot}683$	$B_2 = 2{\cdot}4$ cm. $\pm 0{\cdot}636$
$A_3 = 1{\cdot}6$ cm. $\pm 0{\cdot}690$	$B_3 = 2{\cdot}3$ cm. $\pm 0{\cdot}612$

While the first factor in our analysis simply expresses the fact that the susceptibility to the optical illusion which is being investigated tends to remain a constant quantity for each subject, as compared with the other subjects, the second factor seems to indicate that certain subjects are more susceptible to the illusion under condition A, while other subjects are more susceptible to the illusion under condition B. This finding raises interesting questions regarding the methodology of the psycho-physical methods generally. What precisely are the factors which make condition A different from condition B? How far are factors of this kind general, and how far are they specific to the present test? In the absence of definite experimental evidence, we can here only raise these questions; further research is needed to provide the answers. It may be suggested, however, that a systematic reworking of the standard psycho-physical methods of experiment by means of modern statistical methods of analysis may yield interesting results, and clarify many of the problems left unsolved by the older methods.

On the average, the strength of the illusion was about 25 %, i.e. it was about as strong as when observed by Heymans' subjects under optimum conditions(7). The exact amount of error in the six trials, together with the standard deviations, is given above in Table 2.

It will be noted that with each succeeding trial a practice effect is observed; this practice effect is equally as strong in condition *A* as in condition *B*. This decrease in susceptibility to the illusion with successive trials, in the absence of any information regarding the nature or amount of error to the subjects, is well in line with Judd's results(8). There also appears a tendency, though not as strongly marked, for the standard errors to increase from the first to the third trial in each of the two conditions.

(2) *Correlation between susceptibility to Muller-Lyer illusion and tests of suggestibility*

The correlations between susceptibility to the Muller-Lyer illusion, and the results of eight suggestibility tests, are throughout small and non-significant. The battery of suggestibility tests used has been described elsewhere in detail by the writer(6). In that paper, it had been shown that the customary division into personal, prestige suggestibility on the one hand, and impersonal, non-prestige suggestibility on the other, contradicts the experimental findings, and the results of a factorial analysis carried out on the intercorrelations between the tests. The results of this factorial analysis made it necessary to assume the existence of two entirely unrelated kinds of suggestibility: 'Primary suggestibility', as shown in the Hull Body Sway Test, two Arm Levitation Tests, and the Chevreul Pendulum Test; and 'Secondary suggestibility', as shown in the Progressive Weights and the Progressive Lines Tests, both in their usual form as tests of impersonal suggestibility, and in a modified form, scored as tests of personal suggestibility.

The correlation between primary suggestibility, as measured by the tests mentioned, and susceptibility to the Muller-Lyer illusion, is on the average −0·009. The correlation between secondary suggestibility, as measured by the tests mentioned, and susceptibility to the Muller-Lyer illusion, is on the average 0·007. The detailed correlations are given in Table 3.

Table 3

	Correlation with susceptibility to illusion
Tests of primary suggestibility:	
(1) Body Sway Test	0·044
(2) Arm Levitation, Up	−0·083
(3) Arm Levitation, Down	−0·116
(4) Chevreul Pendulum	0·120
Tests of secondary suggestibility:	
(1) Progressive Weights, Personal	−0·103
(2) Progressive Weights, Impersonal	0·036
(3) Progressive Lines, Personal	−0·054
(4) Progressive Lines, Impersonal	0·149

These correlations disprove rather conclusively the assumption that susceptibility to perceptual illusions is closely connected with suggestibility, as ordinarily understood and used in psychological testing. In fact, susceptibility to the Muller-Lyer illusion fails to show any tendency to correlate either with primary or with secondary suggestibility. We must conclude, then, either that there is no connexion between 'suggestibility' and susceptibility to perceptual illusions, or that such susceptibility constitutes a tertiary kind of suggestibility. It would not seem advisable to make any decision between the two alternatives until more is known about the correlations between susceptibility to one kind of illusion and susceptibility to others.

(3) *Correlation of Muller-Lyer illusion with other data*

Three further sets of data were obtained from the sixty subjects taking part in this study, viz. their intelligence scores on a perceptual test of intelligence, their scores on two perseveration tests, and their scores on two personal tempo tests. Details regarding the actual tests used are given elsewhere(9). These scores were correlated with susceptibility to the illusion, and the following results obtained:

Table 4

	r
Susceptibility to illusion and intelligence	0·213
Susceptibility to illusion and perseveration	0·305
Susceptibility to illusion and personal tempo	−0·196

Of these three correlations, only the one with perseveration is significant, although the other two may be considered suggestive. It has not been found possible to fit these results into any theoretical framework.

It has often been maintained, more particularly by the French school, that there is a particularly close relation between suggestibility and hysteria. In a previous study, the writer failed to find any positive connexion between hysteria and either primary or secondary suggestibility(9). No such connexion between hysteria and susceptibility to the Muller-Lyer illusion became apparent in the present study either. The average error for the hysterics was 1·9 cm., as compared to an average error of 2·2 cm. for the non-hysterics. In view of the high probable errors, this difference is of course insignificant, and its direction would in any case not favour those who might believe that hysteria and (tertiary) suggestibility are positively related.

IV. SUMMARY AND CONCLUSIONS

A test of susceptibility to the Muller-Lyer illusion was given to sixty patients at Mill Hill Emergency Hospital, half of whom were male, half female. The same patients were also given eight tests of suggestibility, a test of intelligence, two tests of perseveration, and two tests of personal tempo. In addition, a clinical diagnosis rated half the patients as (conversion) hysterics, and the other half as free from hysterical symptoms. Correlations were given showing the relationships obtaining between susceptibility to the illusion and the other variables. It appeared that susceptibility to the Muller-Lyer illusion:

(1) Is unrelated to primary suggestibility.

(2) Is unrelated to secondary suggestibility.

(3) Is unrelated to hysteria.

(4) Is related positively to perseveration.

(5) Appears to be related positively to intelligence, although the correlation is not significant.

(6) Appears to be related negatively to personal tempo, although the correlation is not significant.

(7) Remains rather stable from trial to trial.

(8) Depends to some extent on the exact method of measurement adopted.

(9) Shows the influence of training, even in the absence of instruction.

(10) Appears to increase in variability with successive trials.

REFERENCES

(1) Gilbert, J. A. (1894). "Researches on the mental and physical development of school children." *Stud. Yale Psychol. Lab.* II, 40–100.
(2) Seashore, C. E. (1895). "Measurements of illusions and hallucinations in normal life." *Stud. Yale Psychol. Lab.* II, 1–67.
(3) Brown, W. (1916). "Individual and sex differences in suggestibility." *Univ. Calif. Publ. Psychol.* II, 291–430.
(4) Hollingworth, H. L. (1931). "Diurnal variations in suggestibility." *J. Appl. Psychol.* XV, 431–5.
(5) Eysenck, H. J. (1939). "The validity of judgments as a function of the number of judges." *J. Exp. Psychol.* XXV, 650–4.
(6) Eysenck, H. J. (1941). "The validity and reliability of group judgments." *J. Exp. Psychol.* XXIX, 427–34.
(7) Heymans, G. (1896). "Quantitative Untersuchungen über das 'optische Paradoxon'." *Z. Psychol.* IX, 221–55.
(8) Judd, C. H. (1902). "Practice and its effects on the perception of illusions." *Psychol. Rev.* IX, 27–39.
(9) Eysenck, H. J. (1943). "Suggestibility and hysteria." *J. Neurol. Psychiat.* (in the Press).

(*Manuscript received* 2 *February* 1943)

PUBLICATIONS RECENTLY RECEIVED

Straight Thinking in War-Time. By R. H. Thouless. Bickley, Kent: The English Universities Press, Ltd. 1942. Pp. 224. 7*s.* 6*d.* net.

This is a grand book to read, lucid, interesting, extremely clever in a dialectical sense, superlatively honest and a nice, though sometimes rather difficult blend of logic and psychology. It is true that the term "straight thinking" is, or may be, itself not altogether immune from what Dr Thouless calls "incantation value". For it can mean all sorts of different things. It might indicate a type of thinking which goes by the quickest possible route from data to conclusions, or which is ethically right, or which is logically correct, or which is approved because it is what people call "an honest attempt"; and these could get mixed up because, possessing one of these qualities, it might have assigned to it, more or less unconsciously, some of the others. However, it does not seem as if any one of these, or any combination of them, is just what is meant. Many times it is either stated or implied that "straight thinking" and "clear thinking" are the same thing, and if this is interpreted psychologically it perhaps means that anybody thinks straight who so states his thinking and its results that any other reasonably intelligent person can, if he makes a fair effort, tell what facts and principles are being used and how they are being used. In this sense, of course, straight thinking may arrive at conclusions which are not, objectively, in accord with actual facts; and this is admitted many times in the course of the book.

But there is another, and much more troublesome implication which seems to be always creeping in. Dr Thouless believes that "emotion has in our lives the function of driving us to action; absence of emotion, on the other hand, is the condition most favourable to clear thinking". This, however, is consistent with the criterion already suggested only if emotion is coming in unknown to the thinker and unnoticed by his audience. It also suggests, though it may be that this is to push the matter too far, that straight thinking is thinking which has nothing whatever to do with action.

Hidden in this there is a deep and hard psychological puzzle which is likely to bother the psychologically minded reader all through this book. If we assume that somehow clarity of thinking in all circumstances depends upon the possibility of precisely formulating everything that enters into the process, we may be most likely to look for our model among the most objective types of scientific experiment that we can find. For it is there that data, processes and conclusions can all be exhibited in the most satisfying and least ambiguous language possible. Then we may turn to social thinking, or thinking about hopes, fears, desires, ideals and human behaviour generally, and say that this must be cast in the same mould. This is not immediately obvious, however, for the data here are such that perhaps they cannot be fully understood by a process of external inspection alone. It is at least not certainly silly to say that to know how to use them, or to know how others use them, there has to be some affective element in our thinking. And then possibly whether we can say clearly just what this affective determinant is and how it is used, is less important for clarity, than that tolerance should remain when the thinking process has reached some stopping-place for the time being, and that we should recognize that social thinking is not simply *about* social affairs but is itself a social achievement, a synthesis of many individual efforts. There seems to be no way of settling this by general argument. Whether social thinking not only generally does, but in fact must, involve some emotional factors, and can do this without any destruction of intellectual integrity, must await the development of new and better controlled methods of empirical experiment and observation.

If it is admitted as a possibility that social thinking may be a social achievement, requiring a lot of individual efforts but not, in its upshot, fully expressed in any one of them, and certainly not fully expressed by any list of all of them with the items remaining separate, we are up against another difficulty which will bother some psychologists, at least, all through this book. Whenever Dr Thouless has to consider a group "as a unit" he appears to adopt the view that it is nothing but an "organization", with no functions in its own right, so to speak. It is never quite clear how this organization is to be defined, but it appears to be described solely in terms of the area, geographical or human, over which certain customs or rules or laws or institutions are current and common. Again, the manner in which this organization is achieved and maintained does not get any detailed discussion, but it seems to be considered as the work of a leader, a small group of leaders, a minority, or a majority of leading members who, if there are more than one, all think and act uniformly. So if any statement is made attributing some positive achievement to a group, it must be understood to apply only to some member, or at most, to some members of that group. The rest, though perhaps 'in' the achievement are literally not 'of' it.

The difficulty is at its most severe in what is in many ways the most brilliant dialectical chapter of the book, the one called "People, Nations and Rulers". Here, to illustrate his point, Thouless takes eleven different sentences, all purporting to make statements about Germany, and by a discussion which will give any keen reader great intellectual pleasure, claims that these sentences contain at least ten distinct meanings of Germany and that these meanings are likely to get extremely muddled up one with another. Among the eleven statements is one by Goebbels: "Germany is fighting for its existence". Here Germany means "the organization as a nation", and that as "a national unit" it would cease to exist if, for example, it were split up into smaller organizations or taken into a larger one. "This may or may not be true, but

it undoubtedly makes sense." In the same sentence, however, Germany is referred to twice, and in the second case—"its existence"—it is not the national unit which is meant, but only those of the German people who are in the German navy, army or air force, for "what fights is not the organization as a nation, but some of the German people".

Or again, it is often said that "Germany has been planning a war of revenge since 1918"—another of Thouless's eleven sentences. This must, it is maintained, mean only that "some or all of the German leaders have been planning a war of revenge since 1918. The word 'Germany' is here used to refer to the German rulers." To Dr Thouless it appears obvious that "only an individual can make a plan". Apart from the fact that 'planning' in this sentence very definitely includes doing something about a plan when it has been made, the only thing that seems quite clear is that an individual can talk about a plan; who, or what, can make one is a more difficult question, and it may be over-bold to say that a decisive answer is possible as yet.

It may be ethically disturbing, but it seems to be psychologically a fact, that in human society there are all sorts of groups membership of which, even if it is compulsory, carries with it an identification with particular ways of life and of thought. Most people would agree that the family is, or has been, such a group and gains the major part of its strength from that fact. It is not impossible that nations may be too.

There is a chapter on Hitler as a man in which Thouless shows, as Dr William Brown in his well-known lectures on *War and Peace* has shown before him, that Hitler is paranoid or perhaps paranoiac. But though this is true it does not seem to go far enough. The remarkable thing is not Hitler's character, but Hitler's power, and to the making of this have gone plans, organization, ideas and actions a great many of which are far from paranoid. Not only so, but anybody who sets out to trace all these, in their effective forms, to particular individuals and groups of individuals will have his work cut out. Dr Thouless even surmises that very large numbers of the German people are subjected to a kind of *folie à deux* on a grand scale, and will quickly switch from crooked to straight thinking once they are removed from Hitler's direct influence. We may all hope that he is right, but it is not the only psychological interpretation possible.

Supposing, however, that we do learn to think straight about everything, what then? It is abundantly clear, both from his criticisms of recent and current events, and from his exhortations about the future, that Dr Thouless wants action. Are we then to conclude that straight thinking has got after all some kind of an active drive in it, or, having expelled emotion to think straight, must we bring it back to act? Occasional remarks suggest that the only things that really remain very clear as to the correct objects of emotion are objective 'goods' and 'evils'. Different people may become the victims of the one or of the other. Then we hate the evil—war, or 'Hitlerism', or urge to dominance—and we do something about it, but we pity the victim and do nothing about him. Or we love the good—peace, education, straight thinking—and actively promulgate it, but we only feel kindly about the victim and do nothing about him. This is less satisfying than the social psychologist might desire.

There are difficulties in *Straight Thinking in War-Time*: a tendency to underestimate the intellectual respectability of emotion, to assume that straight thinking about society means exactly the same as straight thinking in the case of objective experiment, to suppose that every positive social achievement of a mental order must be traced to that member or those individual members of the group who openly and admittedly contribute something to it which is formulated. But there is a very great amount more, not mentioned in this review, which should be as widely read and as carefully pondered as possible. If the book has even half the success it deserves, its publishers will speedily be putting further editions through the press. F. C. B.

An Introduction to Industrial Psychology. By May Smith. London: Cassell and Co., Ltd. 1943. Pp. 264. 7*s.* 6*d.* net.

This book is essentially what it purports to be, namely, an introduction to and not a text-book of industrial psychology. It introduces the reader to a broad concept of the subject but it does not attempt to explain in detail its various techniques. The author keeps in close contact with the subject-matter throughout by constant reference to original work. On the other hand the book is not a mere catalogue of work done. Original contributions are put in their context and brought into relationship with the whole. The whole for Dr Smith is not the summation of a set of isolated facts but the total integrated personality of the worker.

It is in this respect that this book differs from certain others that have been written on the same subject. From some of these one gets the idea that provided certain selection tests are carried out, and appropriate regulations made about heating, lighting and hours of work, all will be well with the industrial worker. The truth is far otherwise. There is evidence to show that if these matters are not considered the results will be unsatisfactory, but there is no evidence to show that if they are good results will invariably follow. Other essential factors may be lacking which prevents their operation.

There is a curious parallel between industrial psychology and scientific agriculture. The soil chemist can tell us the appropriate fertilizers for different crops, but he cannot guarantee that if these are used the crop will be successful. Satisfactory germination is a biological response to a total environment and so are the reactions of the industrial worker. Unless the whole environment is satisfactory little response will be forthcoming by improving certain aspects of it. The importance of environment is stressed throughout this book, but it is not narrowly defined as consisting mainly of the physical conditions of employ-

ment. It is recognized that one of the most important aspects of environment is the set of human relationships engendered by the interaction of groups and the attitude of those in authority.

The author illustrates this by referring to the fact that sometimes vocational tests fail to give the same results in different settings, although the occupations are identical. From this the conclusion cannot be drawn that tests sometimes measure one thing and sometimes another. Tests never measure vocational success, but only certain capacities shown to be involved in certain skills under certain circumstances. If the circumstances alter there is no reason for supposing that variations in these capacities will be related to variations in skills with which they are associated. However good selection methods are, no good will come from them if training is inadequate or the proper incentives lacking. This aspect of the matter is sometimes overlooked when judging the value of psychological methods in industry, but those who understand this book aright will not be in danger of committing this error, for it is made clear throughout that it is the total environment that matters and not particular aspects of it.

It is the realization of this that leads Dr Smith to pay far more attention to temperamental factors in management than is customary in most books on industrial psychology. No one could draw the conclusion from this book that industrial psychology is a tool in the hands of managements to get the most out of its workers. Instances are given of how certain methods of management adversely affect output and absenteeism, which is important at the present time when so much attention is focused on these matters. It is evident that our efforts to deal with them would be more successful if we sought a remedy not merely by studying the response of the worker to his environment but if we studied the environment itself and those responsible for it.

This book is a short one but none of the space is wasted. It is full of well authenticated scientific conclusions and wise comment on them. Throughout the many years Dr Smith has spent on field investigations for the Medical Research Council's Industrial Health Research Board she has adhered to a rigid scientific technique which makes her critical and fully aware of the limitations of certain results. Her scientific bias does not cloud her sense of values. For her industrial psychology is the study of the whole individual in a particular social group and not the manipulation of a set of measurements. E.F.

Sensation and Perception in the History of Experimental Psychology. By Edwin G. Boring. New York: D. Appleton-Century Co. 1942. Pp. xv + 644.

Prof. Boring's admirable *History of Experimental Psychology* has won an affectionate place in the regard of every serious student of psychology since its publication in 1929. The present volume follows closely in the tradition of its predecessor. Its scope is of course more restricted, and Prof. Boring's retrospective eye this time focuses on problems rather than personalities and on experiments rather than schools. But in the scholarship displayed, and in the attractiveness of the presentation, admirers of Prof. Boring's earlier historical venture will not be disappointed.

The major portion of the text comprises six chapters on vision, three on hearing, one on taste and smell, one on tactual and one on organic sensibility, and a penultimate chapter on the perception of movement and time. The first two chapters, covering the history and physiology of sensation in general, are somewhat sketchy and decidedly inferior to the rest of the book. But the author himself tells us in his preface that these two chapters are less thorough than the rest and intended only as an introduction to what follows. The last chapter, in which Prof. Boring expresses his views on progress in science generally, might fittingly have been relegated to an appendix or even omitted. The important section of the book, however, comprises the thirteen chapters enumerated above, and it is on these that a proper appraisal of Prof. Boring's important book must be based.

The six chapters on vision cover all the principal aspects of visual sensibility and perception as dealt with by the laboratory psychologist. Colour vision, adaptation, space and distance perception, and many other related experimental topics, are surveyed with characteristic thoroughness. Prof. Boring is throughout constrained to preserve his historical perspective, presenting us thereby with a 'longitudinal section' of the whole field of research in contrast to the 'transverse sections' given by the standard text-books. This approach enables us to understand the development of much research whose origin and rationale otherwise appear obscure. Historical accuracy, on the other hand, demands the inclusion of much material whose relevance to the problems of vision in their modern form is admittedly slight. It follows that the student must be advised to read these chapters along with, and not in place of, the standard texts on vision. But this is doubtless Prof. Boring's own intention.

The chapters on hearing comprise two devoted to the more empirical side of the subject—psychophysics of tone and the general phenomena of auditory perception—and one on auditory theory. Taken together, they form a commendable survey of a field in which extensive recent development renders proper historical perspective difficult. The psychology of taste and smell has clearly a limited history, but the author tells us all that can be told. Passing on to the chapter on tactual sensibility, we find the familiar blend of histology, psychophysics and *Beschreibung* that has come to be accepted as experimental psychology. This chapter is well done, and it gives useful historical data on the two-point limen, the Weber function, and related measurements. But it is all rather remote from modern physiology, and reads less like a history than an obituary. (One must hasten to add that this is in no way the fault of the historian!) In the chapter on organic sensibility, on the other hand, the author makes a real attempt to link up the more purely psychological work with physiological studies of visceral sensory patterns

and their role in behaviour. Lastly, we have an excellent résumé of experiments on the perception of time, rhythm and movement. This chapter may do much to re-arouse interest in these important problems. The time is surely ripe for studies of temporal perception which go beyond mere disproof of Vierordt's ill-fated indifference interval.

Prof. Boring has surveyed with considerable success an important province in the development of experimental psychology. His book will be of interest and value to psychologists, sensory physiologists, and especially to students in both psychology and physiology. Yet, after reading his book, one is left with an uneasy feeling that all is not well with the study of perception in its more recent phases. One is aware to-day of a strong tendency to develop the study of sensory processes more and more along the lines of physiological technique and with less and less reference to the established findings of experimental psychology. Psychology and physiology appear to diverge rather than converge. On the one hand we have a considerable body of well-attested facts of experience, especially in the sphere of visual perception, for which explanation is sought by some in the bogus physiology of isomorphism. On the other hand we have accredited theories of vision, based on sound physiological research, whose capacity to explain the most obvious qualities of direct experience is limited in the extreme. This discrepancy is very evident if we compare, say, Koffka's *Principles of Gestalt Psychology* with Rawdon-Smith's *Theories of Sensation*. One may hope that Prof. Boring's book, no less than the research on perception he has himself inspired, will contribute towards a closer liaison between physiological and psychological work on the senses. It may even be that in dedicating his book to Helmholtz the author had this aim in view.

Prof. Boring holds out the hope of a new book devoted to the history of experiment in the sphere of feeling, thought and action. It is earnestly to be hoped that he will see his way to writing this book and thereby completing the historical trilogy. O. L. Z.

Mind: Perception and Thought in their Constructive Aspects. By Paul Schilder. New York: Columbia University Press (London: Humphrey Milford). 1942. Pp. xii+432. 33*s.* 6*d.* net.

The object of this book is to show how the more recent work on the cognitive processes, especially that carried out from the Gestalt standpoint, can be brought into line with the data gained from clinical psychology, and particularly from psychoanalysis. The book contains a great deal of interesting data on the more recondite aspects of perception and thought, such as appear when the normal functioning of these processes is interfered with, either by environmental conditions or when the organism is affected by drugs or disease. The author wishes to trace perception and its relation to feeling to its most primitive and undeveloped form. He considers that there are intermodal connexions and parallelisms between the senses, yet nevertheless the latter have never been completely undifferentiated. But perceived motion and perceived time, and the characteristic rhythms of motion, seem to afford a fundamental intersensory element; and underlie the primitive sensory-motor-'vegetative' (or 'feeling') unit of attitude.

The subject-matter of this book is of great importance; any attempt to link up the data on cognition gathered by experimental psychologists with the observations of feeling and behaviour made by clinicians is valuable. Unfortunately the material itself is exceedingly intractable, and difficult to fit into any coherent pattern. Moreover, the author's style of writing is disjointed, and his habit of leaping rapidly from one aspect of the problem to another, omitting the intermediate connecting links, makes the subject even more difficult to grasp. Thus one is compelled to say that he has made a useful attempt to tackle a fundamental problem, but to regret that he has not further elucidated it.

War over West Ham. By E. Doreen Idle. A Report prepared for the Fabian Society and the Ethical Union. London: Faber and Faber, Ltd. 1943. Pp. 136. 6*s.* net.

Social psychologists have largely failed to grasp the opportunity which the impact of war on our society has given them. Apart from certain limited studies of morale, evacuation, and the effects of the blitz—studies which are suggestive rather than conclusive—psychologists have fought shy of investigating any of the many problems raised. Consequently, we must be grateful for a serious, empirical consideration of these problems, even though this consideration may lack special psychological reference. Miss Idle's book, although primarily sociological in scope, contains much that is important to the social psychologist.

Miss Idle selected a badly blitzed, comparatively self-contained London borough, and investigated on the spot in great detail the effects of the blitz on the people who lived in that borough, on the health, education, recreation and transport services, on the social structure and the whole way of life as it existed there before the blitz. Her examination is remarkably thorough, tracing back the roots of such developments as occurred to their historical foundations in the last century. Her comments are usually fair and unbiased, and while they are restricted to one borough, will probably be found to apply with almost equal force to most other badly blitzed areas.

One shudders to think what would have happened if Mass Observation had taken this subject in the stride of its amiable meanderings, and one feels grateful to the Fabian Society and the Ethical Union for sponsoring this project. One particular virtue of the book in the eyes of social psychologists will be the

fact that it raises a number of problems which appear to be amenable to experimental research: at random we may mention the question of age and adjustment to new situations, the problem of leadership (in shelters, settlements, etc.), and the problem of social coherence (the family group, the street group, the ward, the borough). In the absence of experimental work on these points, we must rest content with the 'case notes' so skilfully presented by Miss Idle. H. J. E.

Nursing Life and Discipline. By SHEILA M. BEVINGTON. London: H. K. Lewis and Co., Ltd. 1943. Pp. xii+89. 7*s*. 6*d*. net.

This monograph describes the results of an inquiry into the conditions of work in hospital nursing as they actually affected the nurses themselves. Dr Bevington carried out confidential personal interviews with each of 525 nurses, comprising 60 % of the probationers, staff nurses and ward sisters in five hospitals, large, medium and small in size, situated in London and the Home Counties (the work was done under peace-time conditions). The main object was to obtain data on the training, discipline and staff relationships in the hospitals, and on the reactions of the nurses to these. Even though the conditions in these hospitals were as favourable as those in any part of the country (it is somewhat a defect of the enquiry that no north of England hospitals were studied), it was obvious that there was much room for improvement. Discipline was often authoritarian and arbitrary, and did not follow the modern conception that the causes of mistakes should be investigated and the offender dealt with in accordance with the particular nature of her case. Thus it compared unfavourably with the discipline of schools and factories, which is also discussed by the author. Moreover, ward sisters were often too rushed and fatigued to be able to devote much attention to the adequate training of probationers. Dr Bevington makes several practical suggestions for the improvement of conditions, including the instruction of senior nurses in training methods, and the appointment of a Welfare Supervisor, similar to factory Welfare Supervisors, to help probationers in their adjustment to hospital life and to supervise their physical, social and cultural needs. The necessity of some such reform is shown by the statement quoted from the Report of the *Lancet* Committee on Nursing that something like 50 % of probationers fail to complete their course.

Dr Bevington must be congratulated on a valuable piece of work which will supplement and underline the conclusions of the Rushcliffe Committee. The psychologist might wish for a rather fuller treatment of the psychological nature of staff relationships, throwing more light on the underlying basis of the strange hierarchy of social attitudes which obtains in this remarkable profession—rather as T. N. Whitehead has described the social relationships of the factory in his *Leadership in a Free Society*. Perhaps only someone who has personally experienced these relationships could fully expound their psychological basis.

Teaching the Individual. By RUTH L. MUNROE. New York: Columbia University Press (London: Humphrey Milford). 1942. Pp. xi+353. 20*s*. net.

This book constitutes a further study in the series devoted to the methods of 'individual education' in use at the Sarah Lawrence College, New York. It describes the research work which has been carried out into the individual characteristics and the psychological problems of the students, and how these affect their attitudes towards their work and the manner in which they perform it. It is shown that some consideration of the student's basic needs, those operating both in her college career and outside it, will give a clearer understanding of her reactions, and also of the methods of study and discipline which will be most profitable for her. Two particular 'educational syndromes' or patterns of behaviour in regard to college work are chosen for detailed studies: that of the 'rigid' or over-conscientious student, unable to show initiative or originality because she cannot 'let herself go' for fear of her own spontaneous impulses; and that of the 'scattered' or undisciplined student, eager and talented, but erratic and incapable of sustained effort because her emotions are always running away with her.

It is recognized that it is impossible to make detailed individual psychiatric studies of every student and then adapt her course of study exactly to her supposed needs—even were such a procedure desirable. Nevertheless, the book ably suggests where for considerable numbers of students 'the shoe pinches'; and how flexibility of educational curriculum and method may be of real value in giving students opportunities of developing in accordance with certain needs which seem to be fundamental for many of them, and of gaining the greatest advantage from their college careers. What is particularly interesting to the psychologist is the suggestion that in such a milieu it is possible for him, after preliminary studies of the students' characters and needs, to predict the students' future development vis-à-vis particular types of curriculum and discipline, and to observe how far his predictions are fulfilled. The studies recorded in this book are admittedly in a preliminary stage; but they are suggestive of very valuable work in future.

Education for Citizen Responsibilities. Edited by FRANKLIN L. BURDETTE. Princeton University Press (London: Humphrey Milford). 1942. Pp. xiii+126. 10*s*. net.

This book contains a series of separate chapters by different authors dealing with the part which has been played, and which might be played, in education for civic responsibility by a study of the following subjects: Anthropology, Economics, Geography, History, Philosophy, Political Science, Psychology, Sociology. The authors are all American, and what they say refers to the American educational system, which differs

in various respects from the English. In particular, there appears to be little differentiation between what is considered suitable for the school curriculum, and what for the University curriculum. There is no line drawn between the theoretical generalizing studies which are within the powers of the young adult; and the mainly factual studies which are more suitable for the adolescent. Doubtless the former might benefit greatly, in regard to education for citizenship, by a study of any of the subjects listed; and in the main the authors give useful suggestions as to how this might be done. The adolescent, if required to study, for instance, Philosophy, Economics or Political Economy, is liable to stray into vague and hazy generalization, and to accept as proven what is still largely hypothetical.

The study of Psychology presents further difficulties. It may be that certain special topics, such as Economical Methods of Study, can be tackled advantageously by older school children. But most psychological topics will, at that age, either appear hopelessly remote, or come so near as to produce introspective self-analysis, and even perhaps a final state like that of the proverbial centipede! This is the danger lying in a fearsome subject called Personality Adjustment (or Improvement) which is mentioned by Prof. Calvin P. Stone, in his chapter on Psychology, as a frequent and desirable feature of the 'high school', as well as the University, curriculum. For children of school age, an able teacher can best begin the study of Psychology through a consideration of the psychological aspects of History, Sociology and Sociological Geography; through them he can inculcate an objective, scientific attitude and a realization of individual differences in character, taste and opinion. M. D. V.

The Psychology of Fascism. By PETER NATHAN. London: Faber and Faber Ltd. 1943. Pp. 158. 8*s.* 6*d.* net.

The argument of this book appears to be that many of the author's patients (who, according to the dedication, have "taught me all I know about the relationships between human beings") developed an exaggerated pseudo-masculinity as the result of their repression of the Oedipus wish and castration fears, and their subsequent identification with the father. This pseudo-masculinity, its admiration of tyrannical authority and its homosexual tendencies, seem often to occur among Fascists; and the author therefore assumes it to be the cause of Fascism. The Germans, it seems, naturally love authoritarianism—the 'leader-principle'—and hate individual responsibility; and therefore take readily to Fascism. But why should they especially among European nations seek this neurotic solution to their infantile wishes and repressions? Even if, as the author seems to believe, the Christian, the socialist and the pacifist, with their repressed aggressiveness and cruelty, are really worse fellows in the long run, still their behaviour *is* different from that of the Fascist. The book seems to provide no explanation of this somewhat fundamental divergence between national groups.

It may be that there are interesting and valuable ideas embedded in the tissue of loose thinking and exaggerated generalization of which this book is made up: they are difficult to extract.

Apparitions. By G. N. M. TYRRELL. Seventh F. W. H. Myers Memorial Lecture. London: Society for Psychical Research. 1942. Pp. 123. 3*s.* 6*d.* net.

The subject of apparitions engaged the attention of the S.P.R. from its earliest days. Much material has been collected and the time was ripe for this judicial and illuminating survey of the data. Mr Tyrrell shows that well-authenticated narratives of apparitions have many features in common which in many cases are not the common features of fictitious ghost stories. For example, the apparition never leaves lasting physical traces, such as footprints, writing, or some such object as a dagger or a faded flower, whereas the ghost of fiction generally does. These common features are regarded as a strong argument for a real foundation of accounts of apparitions and enable Mr Tyrrell to build up an illuminating picture of the 'perfect apparition'.

For theoretical illumination, he turns to the work of philosophers on perception and not to that of experimental psychologists. He contends that, however physically convincing an apparition may appear, there is nothing physically there and no consciousness belongs to the apparition. On the other hand, so many apparitions have been seen by two or more people and have shown genuine correspondence with distant events that there is also no ground for saying that the apparition is a mere hallucination, the product of nothing but the mind of the observer. It is within these limits that an adequate theory of apparitions must be sought. R. H. T.